1

National Library
of Canada

Bibliothèque nationale
du Canada

CANADIAN THESES
ON MICROFICHE

THÈSES CANADIENNES
SUR MICROFICHE

52761

0-315-04867-0

NAME OF AUTHOR/NOM DE L'AUTEUR Gerald Cavallaro

TITLE OF THESIS/TITRE DE LA THÈSE Dynamic Decision Making in Psychophysics: Feedback, No-Feedback, and the Time-Order-Error.

UNIVERSITY/UNIVERSITÉ Carleton University

DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE Ph.D.

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ 1981

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE Dr. William Petrusic

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE

Sept 25, 1981

SIGNED/SIGNÉ

Gerald Cavallaro

PERMANENT ADDRESS/RÉSIDENCE FIXE

19 Gould Street

OTTAWA, Ontario

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

DYNAMIC DECISION MAKING IN PSYCHOPHYSICS: FEEDBACK,
NO-FEEDBACK, AND THE TIME-ORDER-ERROR

by
Gerald Cavallaro

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in Carleton University
1981

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

"DYNAMIC DECISION MAKING IN PSYCHOPHYSICS: FEEDBACK,

NO-FEEDBACK, AND THE TIME-ORDER-ERROR"

submitted by Gerald Cavallaro

in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

.....
Thesis Supervisor

.....
Chairman, Department of Psychology

Carleton University

September 1981

ACKNOWLEDGEMENTS

I would like to thank the members of my doctoral committee for their careful consideration of this dissertation. I would also like to thank Professors Richard Dillon, and William Jones for their help and support. I am indebted particularly to my advisor Professor William M. Petrusic whose patience, insight and encouragement made this work possible. Finally I would like to thank Martha for her patience and encouragement.

Support for this work came from a National Science and Engineering Research Council Grant, APA-0238 awarded to William M. Petrusic.

TABLE OF CONTENTS

	page
List of Tables	ii
List of Figures	iv
Abstract	v
Chapter I	1
Introduction	1
Chapter II	5
Additive learning models and TOE	5
Feedback models	5
No-feedback models and TOE	9
Chapter III	20
Experiment 1	20
Method	21
Subjects	21
Apparatus	21
Stimuli	22
Design	22
Procedures	23
Results	24
Gross data features	24
Analysis of Covariance	28
Time series analysis	35
Relative effects of feedback and driftback	45
Chapter IV	48
Experiment 2	48
Method	49
Subjects	49
Procedures	50
Results and discussion	52
Matching hypothesis	52
ROC analysis	55
Drift-back effects	74
Persistence effects	77
Carry-over effects	81
Chapter V	86
Transfer of Feedback and TOE	86
Chapter VI	92
Experiment 3	92
Method	96
Subjects	96
Apparatus	97
Stimuli	97
Procedures	97
Results and discussion	100
Pre-feedback data features	100
Feedback effects and transfer	104
Chapter VII	111
Transfer of feedback between tasks	111
Chapter VIII	115
Experiment 4	115

Method	116
Subjects	116
Apparatus	116
Stimuli	116
Procedures	116
Design	118
Results and discussion	118
Experiment 5	128
Method	128
Subject	128
Procedure	128
Results and discussion	129
Chapter IX	137
General discussion	137
References	141
Reference note	147
Appendix	148

List of Tables

Table	Page
1. Proportions correct responses in session 1 for Experiment 1.....	26
2. Proportions correct responses in session 2 for Experiment 1.....	27
3. Proportions 'first-longer' responses for Experiment 1.....	29
4. Time series matching chi squares for the continuous feedback groups in Experiment 1.....	39
5. Time series matching chi squares for the intermittent feedback groups in Experiment 1.....	40
6. Time series statistics comparing session 1 with session 2 in Experiment 1.....	41
7. Mean autocovariances in Experiment 1.....	44
8. Sequences of conditions in Experiment 2.....	51
9. Basic statistics and matching chi squares for subject KC in Experiment 2.....	53
10. Basic statistics and matching chi squares for subject KQ in Experiment 2.....	54
11. Matching statistics for feedback conditions for subject KC in Experiment 2.....	56
12. Matching statistics for feedback conditions for subject KO in Experiment 2.....	57
13. Statistics comparing feedback conditions for subject KO in Experiment 2.....	75
14. Proportion 'second-longer' responses for the two presentation orders for Cavallaro (Note 1).....	95
15. Proportions of 'second-longer' responses in the pre-feedback phase of Experiment 3.....	101
16. TOE indices for the pre-feedback phase of Experiment 3.....	102
17. d' indices for the pre-feedback phase of Experiment 3.....	103

18. Proportions of 'second-longer' responses for the transfer phase of Experiment 3.....	106
19. TOE indices for the transfer phase of Experiment 3.	107
20. d' indices for the transfer phase of Experiment 3..	108
21. Difference theory d's for Experiment 4.....	120
22. TOE indices for 3-C trials in Experiment 4.....	123
23. TOE indices for S-D trials in Experiment 4.....	125
24. Basic statistics in Experiment 5.....	130
25. Familiarity theory d's for the two presentation orders in Experiment 5.....	133

List of Figures

Figures	Page
1. Proportions of correct responses for continuous feedback groups in Experiment 1.....	31
2. Proportions of correct responses for intermittent feedback groups in Experiment 1.....	32
3. Mean deviations from matching in Experiment 1..	34
4 -13. Confidence rating-based ROC curves for each phase of Experiment 2.....	59-68
14-17. Proportions of 'first-longer responses plotted in the unit square for Experiment 2.....	70-73
18. Deviations from matching as a function of feedback condition for subject K0 in Experiment 2.....	76
19. Deviations from matching for transitions from continuous feedback for subject K0 in Experiment 2.....	78
20. Deviations from matching for transitions from intermittent feedback for subject K0 in Experiment 2.....	79
21. Deviations from matching for transition from no-feedback for subject K0 in Experiment 2.....	80
22. Deviations from matching as a function of feedback cycle for subject K0 in Experiment 2.....	82
23. TOE as a function of feedback session in Experiment 3.....	105
24. Difference theory d's as a function of feedback stage in Experiment 4.....	121
25. TOE indices from 3-C trials as a function of feedback stage in Experiment 4.....	124
26. TOE indices from S-D trials as a function of feedback stage in Experiment 4.....	126

27. Forced choice TOE indices as a function of
sessions in Experiment 5..... 132
28. Familiarity theory TOE indices in
Experiment 5..... 134

Abstract

Five experiments investigated aspects of dynamic decision processes in psychophysics and the distinction between time-order-errors (TOEs) and decision biases. The first two experiments examined dynamics resulting from omitting feedback in a task involving TOEs. In both experiments, intermittent feedback was less effective than continuous feedback in eliminating TOEs and reducing deviations from probability matching. In terms of additive learning models (ALMs), this result rejected the notion of a stationary criterion on no-feedback trials, and suggested that the criterion drifts back towards its prefeedback origin on no-feedback trials. Drift-back was explicitly demonstrated in experiment 2 in which TOEs and larger deviations from probability matching returned when feedback was withdrawn for an entire session. Both experiments also rejected probability matching as a general consequence of continuous feedback, and therefore also rejected Kac's (1962) ALM, since matching failed to occur when the probability of the long-short (L,S) presentation order (q) was .2. The finding was unique since matching failed to occur in this condition for the group as well as for the individual subject performances. However, matching did occur with $q=.8$ and $q=.5$. Since the $q=.8$ and $q=.2$ conditions differ only with respect to the order of presentation of the stimuli, TOE is implicated in the failures to obtain

matching in the present experiments. Adequate models of feedback effects and bias processes in discrimination must account for the effects of TOE in these situations. The models considered here for intermittent feedback and drift-back fail in this respect since they assume Kac's single parameter model for feedback trials. The final three experiments examined the distinction between TOE and decision bias. It was hypothesized that if TOE was a perceptual phenomenon, similar to contrast for example, then TOE should vary inversely with discriminability. A feedback induced bias that eliminated TOE with less discriminable stimulus pairs then should have been too large for more discriminable pairs. Experiment 3 suggested that TOE is a constant with respect to discriminability since feedback that eliminated the TOE with less discriminable pairs also eliminated TOE with the more discriminable pairs, for which feedback was never provided. The results support Luce and Galanter's (1963) characterization of TOE as a constant criterion bias. Experiments 4 and 5 demonstrated that a feedback effect that reduced TOEs in tasks involving directional judgments, transferred to 'same-different' (S-D) judgments of visual extent, reducing the S-D TOEs. The transfer suggests that S-D and directional judgment tasks are mediated by a common sensory continuum, and therefore rejects the notion that S-D judgments are mediated by a familiarity dimension that has no directional information. Since familiarity is rejected for S-D judgments of visual

extent, then the d' sensitivity measure based on familiarity is inappropriate, and therefore, the measurement of TOE by the difference between the d' s for the two presentation orders does not represent a purely perceptual TOE. These final three experiments provide no support for a perceptual basis for the TOE, and suggest that the reasons for a distinction between criterion bias and TOE are fewer.

Chapter I

Introduction

Psychophysicists have recognized for some time that the effects of bias must be removed from performance in order to characterize an observer's discriminative capacity. Current approaches to psychophysical data do so by defining separate parameters of discriminability and bias. Dusoir (1975) has noted that the adequacy of the definitions is indicated by the way in which certain experimental variables influence only the parameter they would be expected to *a priori*. Hence payoffs, presentation probabilities, and feedback, should affect the bias parameter, but not sensitivity; while signal strength and the difference between stimuli in pair comparisons should covary with sensitivity, but not bias. The d' of signal detection theory (SDT) (cf. Green and Swets, 1965) fulfills the requirement for a sensitivity index fairly well, accounting for its popularity. Bias parameters succeed less well, covarying with sensitivity (Dusoir, 1975 & 1980). In addition, sequential dependencies that reflect a changing bias have been noted in psychophysical data as far back as Fechner (1860). Yet, classic SDT predicts the absence of sequential effects since it: 1) views sensory components as continuous random variables with stationary distributions; and 2)

characterizes the decision rule or bias by a fixed criterion or cutoff on the overlapping distributions of signal and noise. Nor have general theories of the etiology of bias fared well. One view, for example, predicts that subjects will use an optimal decision rule to maximize expected payoffs. But, in detection at least, subjects tend to be conservative in the sense that criteria are usually located between the optimal and the unbiased position (i.e. $d'/2$).

In discrimination designs the bias problems are complicated further by the occurrence of constant errors (CEs). When the method of constant stimuli is used, CEs appear as the difference between the point of subjective equality (PSE) and the standard stimulus, and occur with many stimulus attributes in a consistent manner across subjects. In roving standard designs CEs occur as the stimulus in one ordinal or spatial position is over-estimated with respect to the other. The temptation has been to classify these effects as decisional biases (Allan and Christofferson, 1974; Luce and Galanter, 1963; and Wickelgren, 1968), but such biases would have to be quite unlike those found in detection. For example, one CE, named the time-order-error (TOE) by Fechner (1860), occurs when subjects judge which of two sequentially presented stimuli has the greater magnitude. The error is reflected in more frequent correct judgments with one rather than the other presentation order though the two orders are equally likely, suggesting a bias. But, unlike detection, the bias

is certainly not conservative since the difference between the optimal and unbiased criterion locations disappears when the probability of a signal equals the probability of a noise presentation, which suggests a lack of bias. In addition, the TOE is known to vary systematically with the inter-stimulus-interval (ISI) (Jamieson, 1976; Jamieson and Petrusic, 1976a and b) and even to reverse direction as stimulus intensity increases: effects that seem more like perceptual than bias phenomena. In any case, if CEs are biases they are certainly more complex than those found in detection.

This complexity need not be disadvantageous, however, since the TOE responds to trial-by-trial feedback in a manner consistent with recent models for learning in psychophysics (Jamieson, 1976). Importantly, these models (e.g. Kac, 1962) provide potential solutions for the bias problems noted above.

The approach taken here was to examine the dual problems of bias and CEs in discrimination through the acquisition and transfer of feedback induced decision rules. It was assumed that the amount of TOE gauges feedback's effectiveness. This made possible the measurement of the relative effectiveness of different feedback schedules and the effects of removing feedback. The first two experiments concern these measurements and provide new information about the dynamics of the decision process in psychophysics. In addition, using the TOE in this way allowed an assessment of

how well feedback training transfers to new situations. The final three experiments concern such transfer. It was hoped that the transfer paradigm could provide a basis for elaborating the relationship between CEs and bias by: 1) assessing transfer between different judgmental tasks that are potentially mediated by different sensory qualities; and 2) examining the functional relationship between TOEs and the difference between stimuli in a pair on a single stimulus attribute. In following sections, the implications of transfer will be discussed after first considering some of the learning models and their relationship to TOEs.

Chapter II

Additive Learning Models and TOE

Feedback Models

One approach to the difficulties encountered with notions of bias has been the development of learning models applicable to psychophysical data. A number of the models were introduced to account for sequential dependencies (Atkinson, Carterette & Kinchla, 1962, Atkinson & Kinchla, 1965, & Friedman, Carterette, Nakatani & Ahemada, 1968), while others were prompted by the recognition that steady-state psychophysical performance must be the end result of a learning process (eg. Bush, Luce & Rose, 1964, & Kac, 1962). Of the various models only three (Kac, 1962, Sklansky, 1963, & Schoeffler 1965) assume a continuum of sensory states compatible with SDT. Of these only the Kac model and its variations will be considered here since it has received the most mathematical and empirical attention.

Models of the form proposed by Kac are termed additive learning models (ALMs) since trial by trial changes do not depend on the amount of previous learning, and the effects sum algebraically. These models adopt the activation and decisional assumptions of SDT but propose that the decision rule is dynamic. Hence, in a detection task, which requires the subject to respond 'yes' or 'no' to indicate the presence or absence of a signal, the effects of sensory

stimulation are summarized by a continuous random variable X which is compared to a criterion C . If X exceeds C the response is 'yes', otherwise, 'no'. However, C is assumed to change over trials. The earliest ALM (Kac, 1962) assumed that the value C was adjusted by some small amount A whenever the subject made an error, so that if an error occurred on trial n the criterion on trial $n+1$:

$$C_{n+1} = C_n + A \text{ if } X_n > C_n \text{ and}$$

$$C_{n+1} = C_n - A \text{ if } X_n < C_n, \text{ otherwise } C_n.$$

This model has been generalized by Dorfman and Biderman (1970) so that the value of the adjustment might depend on the various kinds of trials and their outcomes. Hence, there are four possible A_{ij} in a signal detection task where $i=1,2$ depending on whether a signal or noise trial occurred and $j=1,2$ depending on whether the response was yes or no. Thomas (1973) has further generalized the model so that: 1) criterion change is probabilistic; 2) expected criterion movement is a function of the distance of the evidence from the criterion (i.e. $X_n - C_n$); and 3) the effects of asymmetric payoff matrices can be incorporated [Dusoir (1980) provides a catalogue of the variations of Kac's model and their predictions]. These models are important because they provide potential solutions, under fairly general conditions, to two problems with the characterizations of bias in psychophysical tasks: sub-optimal bias locations (Thomas, 1973) and sequential dependencies in the data.

Certain of the models predict bias more specifically. Consider the derivation, arrived at independently by Norman (1972) and Thomas (1973), that Kac's single parameter error-correcting model predicts the proportion of responses indicating that a stimulus was presented should equal its probability of presentation (ie. probability matching). Some support for this prediction is provided by Dorfman, Saslow and Simpson (1975) and Jamieson (1976). As well, the estimates of the A_{ij} obtained by Dorfman et al. for the four parameter general model closely approximate the error-correcting assumptions of Kac's single parameter model. On the other hand, data from Dusoir (1980) suggests the general model, or that different subjects are characterized by different models. Reasons for the discrepancies are not clear. Nevertheless, the majority of subjects that have participated in experimental tests of the models (cf. Dorfman et al., 1975 and Dusoir, 1980) have approximated matching reasonably well.

The models are also applicable to the CEs that occur in discrimination designs. In a duration discrimination experiment, for example, the subject judges whether the first stimulus is longer or shorter than the second. Consider the case in which two confusable durations are presented in pairs an equal number of times in each temporal order. Thurstonian representations are usually assumed for the stimuli. Hence, the stimuli, S_1 and S_2 , are represented by random and normally distributed variables, X_1 and X_2 ,

respectively. The subject accomplishes the task by comparing the subjective difference $X_1 - X_2$ with a criterion C . In the unbiased case C is zero. Hence, the subject responds 'first longer' if $X_1 - X_2 > C$, otherwise 'first shorter'. It is easy to see that if the criterion C is not set at zero or if the differences $X_1 - X_2$ are not symmetrically distributed around zero, correct responses will be greater for one of the two presentation orders, resulting in TOEs. However if, as a result of feedback, a subject adjusts C to obtain probability matching, TOE is eliminated since the proportion of correct responses would be the same with each presentation order. Jamieson (1976) eliminated TOE with the provision of feedback when presentation probabilities were equally likely and obtained matching for other presentation order probabilities (experiment 4).

It is interesting to note that the probability matching result has been proposed as a likely consequence of: 1) providing feedback (Norman, 1972), and 2) specifically not providing feedback in a signal detection or recognition task (Thomas and Legge, 1972). Hence, there seems to be no a priori reason to expect asymptotic criterion location to vary with feedback's presence or absence in such tasks. Dorfman et al. showed that subjects did not match early in feedback training, but no experimenters have compared the presence and absence of feedback in the same detection or recognition experiment. Jamieson (1976), however, used a

duration discrimination task that was characterized, in the absence of feedback, by substantial departures from probability matching due to TOEs. The matching obtained by Jamieson may then be attributed uniquely to the effect of feedback.

No-feedback Models and TOE

If feedback processes have received little attention in psychophysics, the effects of not providing feedback have received less. The problem is considered here in the context of: 1) omitting feedback on a random portion of trials, and 2) withdrawing it completely after some amount of feedback training. Given the lack of data, numerous models are possible. Three simple ones receive consideration here without rigorous mathematical treatment since the main purpose is to provide a framework for interpreting possible experimental outcomes and further elaborating the dynamics of psychophysical decision processes. The formulations model two possible effects in discriminations which involve TOE: 1) no effect of removing feedback, and equal effects for continuous and intermittent feedback; and 2) criterion drift-back, and a reduced effect for the intermittent schedule. Two drift-back models are presented, one that suggests a bias explanation for TOE, and one based on a perceptual TOE effect.

Intermittent feedback has been investigated in probability learning tasks which require the subject to predict which of several randomly scheduled stimuli will be presented on a particular trial. Feedback takes the form of simply presenting the scheduled stimulus. In this task too, subjects often probability match (cf. Neimark & Estes, 1967, and Lee, 1971; but see also Norman, 1971 for a different view). Intermittent feedback slows rather than prevents asymptotic matching (Neimark & Estes, 1967), which prompted Greeno (1967) to suggest that the learning parameter for trials without feedback is nil. For model 1, we assume Kac's matching model for feedback trials, and stationary criterion on no-feedback trials. On average then, for a series of m trials, m large, the following expression can be written:

$$C_{n+m} - C_n = AfP(F) [P(R1/S, L)(1-q) - P(R2/L, S)q] \quad (1)$$

where Af is the constant of criterion change occurring on the proportion $P(F)$ of feedback trials; S, L and L, S are the two presentation orders; q is the probability of the L, S order; and $R1$ and $R2$ are the responses indicating which is the larger stimulus.

Note that, if Af is assumed to be constant for all $P(F)$, then the distance the criterion travels in m trials is inversely proportional to the probability of feedback. In

addition, assuming the process reaches asymptote, then on average,

$$C_{n+m} - C_n = 0 \quad (2)$$

Hence,

$$0 = P(R1/S, L) (1-q) - P(R2/L, S)q \quad (3)$$

And by adding $P(R1/L, S)q$ to both sides of equation 3 and rearranging terms,

$$P(R1/S, L) (1-q) + P(R1/L, S)q = P(R1/L, S)q + P(R2/L, S)q \quad (4)$$

which gives the probability matching result,

$$P(R1) = q \quad (5)$$

Hence, intermittent feedback should be asymptotically as effective as continuous feedback, but the rate of approach to matching should be inversely proportional to $P(F)$. Additionally, the withdrawal of feedback should leave asymptotic rates unchanged.

The second formulation also assumes Kac's single parameter model for feedback trials. For no-feedback trials we employ a model investigated by Norman (1972) and examined by Dusoir (1980) in an experiment without feedback. The

no-feedback model assumes two constants of criterion change, $-A_1$ and $+A_2$, which are effective following R_1 and R_2 responses respectively. Norman (1972) has shown that asymptotically, this model implies,

$$P(R_1) = A_2 / (A_1 + A_2) \quad (6)$$

Note that, when $A_1 \neq A_2$, equation 6 describes a dynamic process for TOE. In addition, the process could be described as a bias unless the origin of the unequal A_i is in some way sensory. As it stands the model seems implausible since $P(R_1)$ does not depend on q , though it would if the A_i varied with q .

For intermittent feedback at asymptote, the model is,

$$\begin{aligned} 0 = & P(F)A_f[P(R_2/L, S)q - P(R_1/S, L)(1-q)] \\ & - P(N)[A_1P(R_1) - A_2P(R_2)] \end{aligned} \quad (7)$$

where $P(N)$ is the probability of omitting feedback.

For ease of exposition, assume $P(F) = P(N) = .5$. Using the methods of equation 4 we obtain,

$$0 = A_f[q - P(R_1)] - A_1P(R_1) + A_2P(R_2) \quad (8)$$

and,

$$0 = A_f[q - P(R_1)] - A_1P(R_1) + A_2 - A_2P(R_1) \quad (9)$$

which reduces to

$$A_2/(A_1+A_2) = P(R_1) - A_fq/(A_1+A_2) + A_fP(R_1)/(A_1+A_2) \quad (10)$$

and then,

$$A_2 + A_fq/(A_1+A_2) = P(R_1) [1 + A_f/(A_1+A_2)] \quad (11)$$

and finally,

$$P(R_1) = (A_2 + A_fq)/(A_1 + A_2 + A_f) \quad (12)$$

The expression becomes more meaningful if we scale the feedback parameter, A_f , in terms of the sum of the no-feedback parameters, $A_1 + A_2$. Letting $A_f = k(A_1 + A_2)$ we obtain,

$$P(R_1) = A_2/[(1+k)(A_1+A_2)] + k(A_1+A_2)q/[(1+k)(A_1+A_2)] \quad (13)$$

which reduces to,

$$P(R_1) = P_n(R_1)/(1+k) + qk/(1+k) \quad (14)$$

where $P_n(R_1) = A_2/(A_1+A_2)$ is the probability of response R_1 when no feedback has been provided. With prior knowledge of

$P_n(R_1)$ k can be estimated using equation 14. Note that deviations from matching are an inverse function of k , the effectiveness of feedback, and:

1) when $k=0$,

$$P(R_1) = P_n(R_1) ; \quad (15)$$

2) when $k=1$ (i.e. $A_f=A_1+A_2$),

$$P(R_1) = 1/2 [P_n(R_1) + q] ; \quad (16)$$

and 3)

$$P(R_1) \rightarrow q \text{ as } k \rightarrow \text{infinity}. \quad (17)$$

Hence, unless A_f is very large with respect to the no-feedback bias parameters, intermittent feedback should produce R_1 rates intermediate to continuous and absent feedback. Additionally, when feedback is no longer presented, $P(R_1)$ should drift back to its no-feedback level.

The final model follows Parks' (1966) and Thomas and Legge's (1970) proposal that in the absence of feedback and knowledge of the a priori stimulus probabilities, subjects attempt to match $P(R_1)$ to q . A special case of the matching hypothesis considered by Thomas and Legge (1970) suggests that subjects arrive at a criterion C that enables $P(R_1)$ to match q . Thomas and Legge suggest that in the absence of

knowledge about q , subjects choose a value themselves. How subjects choose q is not specified, though they might take into account the build-up of the distribution of sensory effects over trials.

When TOEs are present it is clear that $P(R1)$ does not match q when q is .5. But q may be chosen incorrectly if the distribution of sensory effects is not symmetrical, or if the subject attempts to use the unbiased criterion $C=0$, initially, but the sensory effects are not distributed symmetrically around zero. Such asymmetries might easily arise from a perceptually based TOE. However q is chosen, the present model assumes that $P_n(R1)$ is the observer's subjective estimate of q , based on a cut-point $C0$, arrived at in the absence of feedback. In addition, it is assumed that $C0$ has attractive properties.

Again the Kac model is assumed for feedback trials. It is further assumed that the observer compares the sensory value $X1-X2$ with the criterion C_n , and also with the memory of $C0$. Again, response $R1$ occurs if $X1-X2 > C_n$, otherwise $R2$. However, if feedback is not given on trial n , the criterion on trial $n+1$ becomes,

$$C_{n+1} = C_n + A_n \text{ if } C_n \leq X1-X2 < C0$$

$$= C_n - A_n \text{ if } C_n > X1-X2 \geq C0$$

$$= C_n \text{ otherwise.} \quad (18)$$

Hence, the model assumes the criterion drifts back towards its origin if C_n and C_0 yield different responses on a particular trial.

Assuming the process reaches asymptote, the following expression can be written,

$$\begin{aligned} 0 = & P(F) A_f [P(R1/S, L) (1-q) - P(R2/L, S) q] \\ & - P(N) A_n [P(R2/Y > C_0) P(Y > C_0) - P(R1/Y < C_0) P(Y < C_0)] \end{aligned} \quad (19)$$

where $Y = X_1 - X_2$. Note that when feedback is given on every trial we obtain

$$P(R1) = q$$

the matching result. When no feedback is provided we obtain the following:

$$0 = P(R1/Y < C_0) P(Y < C_0) - P(R2/Y > C_0) P(Y > C_0) \quad (20)$$

or,

$$\begin{aligned} & P(R2/Y > C_0) P(Y > C_0) + P(R1/Y > C_0) P(Y > C_0) \\ = & P(R1/Y < C_0) P(Y < C_0) + P(R1/Y > C_0) P(Y > C_0) \end{aligned} \quad (21)$$

which reduces to

$$P(Y > C_0) = P(R_1) \quad (22)$$

Since we have assumed that $P(Y > C_0)$ is the same as the observer's subjective estimate of q , and that $P_n(R_1)$ gives that value, then $P(R_1)$ should return to its pre-feedback level after feedback has been withdrawn.

Furthermore, for intermittent feedback, the asymptotic expression becomes

$$0 = P(F)A_f[P(R_1) - q] - P(N)A_n[P(Y > C_0) - P(R_1)] \quad (23)$$

Again assume for convenience that $P(F) = P(N) = .5$, and that $A_f = A_n$ then,

$$0 = P(R_1) - q - P(Y > C_0) + P(R_1) \quad (24)$$

and,

$$P(R_1) = 1/2 [q + P(Y > C_0)] \quad (25)$$

Since it has been assumed that $P_n(R_1)$ is a subjective estimate of q , and $P(Y > C_0)$, then

$$P(R_1) = 1/2 [q + P_n(R_1)] \quad (26)$$

Again we have the result that intermittent feedback should be only partially effective compared to continuous feedback. Its relative effectiveness, of course, would depend on the relative strengths of A_f and A_n . The ratio A_f/A_n can be estimated from equation 23 by substituting $P_n(R1)$ for $P(Y > C0)$, if $P_n(R1)$ has been estimated before feedback is provided. The appropriate equation is

$$A_f/A_n = [P(R1) - P_n(R1)] / [q - P(R1)] \quad (27)$$

which, interestingly enough, is equivalent to the estimate k in model 2 (equation 14).

The models presented in this section do not result in any startling predictions, but they show that drift-back phenomena can be incorporated into ALMs. Whether such elaborations are warranted depends on whether driftback takes place, and if it does, whether plausible models are mathematically tractable and empirically testable. One clear problem is that Kac's (1962) model, assumed here for feedback trials because of its mathematical simplicity, may not be a good characterization of feedback processes. If adoption of the four parameter general model is necessary, as Dusoir (1980) fears, then adding parameters for no-feedback trials further complicates the task of characterizing the decision process. Nevertheless, dynamic consequences of feedback's omission would be interesting, particularly if they systematically undo feedback's effects as models 2 and 3 suggest.

The following two experiments examine the possible dynamic consequences of not presenting feedback. Both concern the effects of intermittent feedback, but the second examines the drift-back hypothesis directly through feedback's complete removal.

Chapter III

Experiment 1

The present experiment was designed to examine the relative effectiveness of intermittent feedback in a duration discrimination task that reliably results in TOEs. To this end, the probability of feedback and the probability of presenting the stimuli in the order $\langle L, S \rangle$ were jointly varied so that the matching hypothesis [i.e., $P(R1) = q$], as well as intermittency, could be evaluated under a number of conditions. In addition, performance was observed before feedback's introduction so that changes in $P(R1)$ could be attributed to the feedback.

The stimulus durations were short (i.e., less than 300 msec) and expected to produce positive TOEs. Hence, over-matching [i.e., $P(R1) > q$] was expected when q was .5. Since such TOEs are fairly ubiquitous, over-matching was expected at the other levels of q as well. Hence, according to the Kac (1962) model, continuous feedback should reduce $P(R1)$ to matching.

Since it seems inconsistent with the flavor of ALMs that feedback should be effective only if it occurs on every trial, three modifications of Kac's single parameter model were presented in the preceding section for intermittent feedback. The first model contrasted with the other two for

asymptotic performance, predicting that intermittent feedback should only slow asymptotic matching, not prevent it, whereas models 2 and 3 predict a reduction but not the complete elimination of over-matching. In addition, two of the models allow estimation of the comparative effects of feedback and no-feedback with prior knowledge of $P(R1)$ in the absence of feedback. Since $P(R1)$ was obtained for each subject before the introduction of feedback, such estimations are attempted for the present experiment.

Method

Subjects

Thirty undergraduate students who were enrolled in an introductory psychology course participated in the experiment. Each subject received course credit and an hourly wage, and in addition, received \$.01 for each correct response and was fined \$.01 for each error. Knowledge of the amount earned was kept secret until the subject completed the experiment. Four additional subjects were replaced, due to zero TOEs in two cases, and small negative TOEs in the others, all when q was .5.

Apparatus

A Data General Nova 1220 computer controlled the presentation of instructions, feedback, and stimuli; event sequencing; and the recording of responses and response times. A response panel, along with an Electrohome video monitor for presenting instructions and feedback, and a Monsanto light-emitting diode (LED) for presenting stimuli

were interfaced with the computer, located in a separate room. The response panel contained two sets of four response keys. Arranged in a cross pattern in the center of the panel were keys marked 'yes', 'no', 'first', and 'second'. The 'first' and 'second' keys were horizontally arranged. The vertically arranged 'yes' and 'no' keys were not used. In a horizontal row at the top of the panel were four keys marked 'guess', 'little certainty', 'moderate certainty' and 'certain' which were used for confidence ratings. The LED was affixed to a transparent, red, plastic screen which was situated behind the response panel and in front of, but below the video monitor.

Stimuli

Two stimulus durations, 250 and 280 msec in length, were defined by the period of illumination of the LED. The durations were timed to within ± 1 msec by a 100 KHz crystal clock.

Design

Subjects were randomly assigned to six groups formed by the factorial combination of three probabilities (q) of the L,S (280,250) presentation order (.2,.5, and .8) and two schedules of feedback (continuous and intermittent). In addition, the experiment was conducted in two sessions with feedback provided only during the second session. Hence, the presence versus absence of feedback was a within subject factor along with presentation order (L,S/S,L) and instructions ('shorter' and 'longer'). Following the first

twenty trials of the second session, continuous feedback subjects received feedback on every trial, intermittent subjects on a random half of the trials.

Procedures

Each subject participated in two, one and one half hour sessions consisting of 420 trials each. The two sessions were separated by at least one hour, but due to scheduling difficulties, as many as seven days intervened between sessions. Subjects were allowed a five min rest after the first 120 trials and then after each subsequent 100 trials. On each trial the subject was required to select the shorter (or longer) of two stimulus durations by pressing one of two response keys, marked 'first' or 'second', and then indicate a level of confidence in the judgment by pressing one of the four confidence keys. A typical trial progressed as follows: 1) an instruction ('shorter -- first or second -- ?' or 'longer -- first or second -- ?') was displayed on the video screen for 2.5 sec, and then removed; 2) following a 1 sec delay the LED was illuminated for one of the two stimulus durations (280 or 250 msec); 3) after a 1 sec ISI the LED was re-illuminated for the period of the other duration; 4) on feedback trials a phrase indicating the correct response appeared on the screen immediately following the subject's confidence response and was displayed for 2.5 sec followed by a 2 sec inter-trial-interval (ITI); on no-feedback trials a 4.5 sec ITI began immediately after the subject's confidence response.

The feedback phrases were displayed in the center of the video screen and they restated the instruction for emphasis. Hence, if the instruction was 'shorter -- first or second -- ?' and the first duration was the shorter, then the feedback was 'first -- shorter'.

Each session consisted of 21 blocks of 20 trial units. Within each unit the proportion of L,S trials was constant for a particular group and each of the two instructions appeared equally often with each of the presentation orders. In addition, for intermittent feedback groups, feedback was provided on half of each of these four combinations of instructions and presentation order. Otherwise the sequence of trials was random within each unit.

Before the experiment began each subject was familiarized with the setting and informed of the wages and payoffs. The subject was then instructed to select the shorter duration on trials beginning with the instruction 'shorter -- first or second -- ?', and the longer duration on trials beginning with the instruction 'longer -- first or second -- ?' by pressing the appropriate key, and then to press one of the confidence keys to indicate subjective confidence in the judgment. This was followed by 20 practice trials.

Results

Gross Data Features

Some basic statistics of experiment 1 are presented in Tables 1, 2, and 3. These tables and subsequent analyses

exclude the first 20 trials, which were used for warm-up. Tables 1 and 2 present each subject's proportions of correct responses $P(c)$ for sessions 1 and 2, respectively. Columns include the $P(c)$ statistic for each order of presentation and collapsed over presentation order. Greater detail for the $P(c)$ measure can be found in Tables A1-A12 of appendix A where the scores are further partitioned by instruction and blocks of 100 trials. That subjects discriminated the durations can be seen in Table 1 where the overall $P(c)$ scores are all above chance, though the difficulty of the discrimination is evidenced by the closeness of many of these scores to .5. The presence of positive TOEs is demonstrated by the better performance of all groups with the L,S presentation order. Unfortunately, the TOEs were not equivalent for comparable feedback groups. When q was .5 the continuous group was clearly superior to its respective intermittent group with the L,S order. This difference, which was clearly due to sampling error, motivated the use of analyses of covariance presented below.

Table 2 shows a slight general improvement in the overall $P(c)$ scores with respect to the first session although subject 18 dropped just below the chance level in session 2. Importantly, for the continuous feedback groups, the superiority of the L,S order is reduced compared to Table 1 when q was .5, reflecting a smaller TOE, but it is increased when q was .8 and largely unchanged when q was .2. In addition, these effects are largely due to changes in the

TABLE 1

Proportions of correct responses with the -long,short and short,long orders and collapsed over presentation order in session 1.

<u>q = .2</u>					<u>q = .5</u>					<u>q = .8</u>				
	<u>Subject</u>	<u>L,S</u>	<u>S,L</u>	<u>Mean</u>		<u>Subject</u>	<u>L,S</u>	<u>S,L</u>	<u>Mean</u>		<u>Subject</u>	<u>L,S</u>	<u>S,L</u>	<u>Mean</u>
Continuous Feedback	1	.58	.60	.60		6	.97	.29	.63		11	.82	.55	.77
	2	.48	.69	.65		7	.83	.26	.55		12	.74	.47	.69
	3	.60	.66	.65		8	.95	.67	.81		13	.65	.54	.63
	4	.80	.53	.58		9	.76	.38	.57		14	.69	.60	.67
	5	.74	.63	.65		10	.79	.47	.63		15	.64	.55	.62
	Mean	.64	.62	.63		Mean	.86	.41	.64		Mean	.71	.54	.68
<hr/>														
Intermittent Feedback	16	.33	.73	.65		21	.75	.45	.60		26	.93	.67	.88
	17	.54	.66	.64		22	.91	.18	.55		27	.75	.53	.71
	18	.66	.49	.52		23	.65	.51	.58		28	.58	.76	.62
	19	.60	.56	.57		24	.64	.59	.62		29	.59	.76	.62
	20	.81	.49	.55		25	.74	.47	.61		30	.71	.44	.66
	Mean	.59	.59	.59		Mean	.74	.44	.59		Mean	.71	.63	.70

P(c) scores with the S,L order, which increase when q was .5 and decrease when q was .8 for each subject. Changes for the intermittent groups appear much less consistent.

The increased superiority of the L,S order for $q=.8$ with continuous feedback is consistent with the matching hypothesis. This can be seen by referring to Table 3 which presents summary P(R1) data for each subject in the two sessions. Again, more detailed P(R1) data can be found in Tables A13-A18 in appendix A where the measure is presented for each instruction in each block of 100 trials. Table 3 shows that, contrary to expectation, all but one subject in the $q=.8$ groups under-matched. This finding is also at variance with Jamieson (1976) who obtained over-matching with $q=.7$. The reasons for this discrepancy are not clear. However, in Jamieson's experiment, unlike here, the subject was informed of q , leading perhaps to a less conservative criterion placement. Nevertheless, because of this under-matching, subjects had to increase responding R1 in order to match when q was .8. Over-matching was obtained in session 1 for the other values of q , and was reduced in session 2 when $q=.5$.

Analyses of Covariance

The apparent differences between intermittent and continuous feedback groups in the first session motivated the use of analysis of covariance to summarize feedback's effects. The proportion correct data in Tables A1-A12 were normalized with an arcsine transformation and the scores

TABLE 3

Proportions of 'first — longer' responses in sessions 1 and 2 of Experiment 1.

q = .2			q = .5			q = .8			
Subject	Session 1	Session 2	Subject	Session 1	Session 2	Subject	Session 1	Session 2	
Continuous Feedback	1	.44	.54	6	.84	.53	11	.75	.84
	2	.34	.37	7	.79	.50	12	.70	.72
	3	.39	.51	8	.64	.61	13	.61	.81
	4	.54	.30	9	.69	.39	14	.63	.77
	5	.44	.31	10	.66	.62	15	.60	.69
	Mean	.43	.40	Mean	.72	.53	Mean	.66	.76
Intermittent Feedback	16	.28	.26	21	.65	.57	26	.81	.88
	17	.38	.47	22	.87	.76	27	.69	.62
	18	.54	.64	23	.57	.74	28	.51	.75
	19	.47	.39	24	.53	.54	29	.52	.70
	20	.57	.44	25	.64	.50	30	.68	.75
	Mean	.45	.44	Mean	.65	.62	Mean	.64	.74

analysed using the last block of session 1 as the covariate for the four blocks of session 2. The design was a split-plot factorial with q and $P(F)$ as between subject; and presentation order, instructions, and blocks of 100 trials as within subject variables. Tests of main effects and lower-order interactions were complicated by two significant four-way interactions: 1) blocks by instructions by $P(F)$ by q ($F=3.06$; $df=6,72$; $p<.05$); and 2) blocks by order by $P(F)$ by q ($F=2.73$; $df=6,72$; $p<.05$). The interaction with instructions appeared idiosyncratic and will not be considered. Figures 1 and 2, which presents the mean $P(c)$ scores with each order in each block of 100 trials for each group, provides a first approximation towards interpreting the interaction with presentation order. These group mean curves suggest that the effects of feedback are negligible when provided intermittently. Continuous feedback, however, results in convergence of the $P(c)$ scores for the two presentation orders when $q=0.5$, but divergence when $q=0.8$.

In order to simplify interpretation the order variable was nullified by calculating a deviation from matching statistic $DM=P(R1)-q$, which reflects the deviation of $P(R1)$ from probability matching. The DM data were then analysed by analysis of covariance, again using the last block of session 1 as the covariate. The rationale for this statistic is simply the matching hypothesis. For the DM variable the highest-order interaction was among the blocks, $P(F)$, and q factors ($F=2.47$; $df=6,72$; $p<.05$). This

Figure 1. Proportions of correct responses for each presentation order during each 100 trial block in the no-feedback and feedback sessions for the three continuous feedback groups.

Figure 2. Proportions of correct responses for each presentation order during each 100 trial block in the no-feedback and feedback sessions for the three intermittent feedback groups.

interaction is depicted in figure 3 which presents the DM statistic collapsed over instructions, as a function of blocks for the six groups. Again, there appears to be virtually no effect of intermittent feedback while the means of the three continuous feedback groups attain or are approaching matching (ie. $DM=0$)

To check on the rate of approach of the three curves toward probability matching the DM scores for $q=0.8$ subjects, who were undermatching in session 1, were multiplied by -1.0 so that the direction of approach was the same for all three groups. The analysis of covariance now revealed a simple blocks by $P(F)$ interaction ($F=5.24$; $df=6,72$; $p<.005$) with a difference in linear trends over blocks for the two feedback conditions ($F=7.08$; $df=1,24$; $p<.05$), suggesting that the effects of continuous feedback on DM were not reliably different for the three q groups.

The results are rather clear concerning the ineffectiveness of intermittent feedback, suggesting that the criterion drifts back on no-feedback trials. Hence, there may be dynamic consequences of omitting feedback that are opposite in direction to the effects of feedback. This possibility will be explored more directly in the next experiment.

Continuous feedback, however, is effective, reducing TOE when $q=.5$ and causing all groups to approach matching $P(R1)$ to q . With respect to matching, the results are at least provocative and suggest testing the hypothesis that

Figure 3. Deviations from matching for the six groups during the last block of 100 trials in the no-feedback session and each block of 100 trials in the feedback session. Points plotted are $P(R1)-q$. Circles, triangles, and diamonds indicate groups for which the proportion of L,S trials (q) were .2, .5, and .8 respectively. Closed symbols indicate continuous, and open symbols indicate intermittent feedback groups.

individual subjects matched $P(R_1)$ to q . This hypothesis is considered next.

Time Series Analysis

The procedures used in this section are intended to permit statistical inference with dependent observations such as might be obtained in a sequence of trials from a single subject. The central idea is that dependencies among trials are measured as covariances and included with the sample variance of the sequence so that the independence assumption is no longer needed. Norman (1971) provides the derivations and assumptions needed to extend classical procedures to this situation, and an example. Excellent summaries and recent applications can be found in Dusoir (1974, 1980). Since Dusoir's applications are directly relevant to the present experiment, the procedures and their explication presented in the following are largely his. Some additional material based on Norman's (1971) paper is also presented.

Consider a series of n trials obtained from a single subject and denote the response on the i th trial $0 < i < n$, by the random variable X_i . Let $X_i = 1$ if the response was 'first longer'; $X_i = 0$ if 'second longer'. We assume that the trial process is second order stationary, such that $E(X_j) = u$ and the autocovariance of lag k , $\text{cov}(X_j, X_{j+k}) = c(k)$ both exist and do not depend on j and that the sum of the absolute values of $c(k)$ is finite. Then let

$$N = n - m + 1, \quad (28)$$

$$\bar{X} = \sum_{j=m}^n X_j / N,$$

$$c(k) = (1/N) \sum_{j=m}^{n-k} (X_j - \bar{X})(X_{j+k} - \bar{X}), \quad (30)$$

$$\text{Var} = c(0) + 2 \left(\sum_{k=1}^K c(k) \right). \quad (31)$$

Then N is the number of trials under consideration, $\bar{X}$ is the sample mean, $c(k)$ is the sample autocovariance of lag k , and Var is the sample variance. Since the idea is to include in Var the $c(k)$ that represent dependencies among the trials, K in (4) is chosen so that $c(k)=0$ for $k>K$.

The following familiar statistics were used by Dusoir (1974, 1980):

$$t = (\bar{X} - u) \sqrt{N} / \sqrt{\hat{\text{Var}}} \quad (32)$$

which is asymptotically normal for large N (Norman, 1971); and

$$\chi^2 = \sum_{j=1}^I (t_j)^2 \quad (33)$$

where t_j denotes the value of t in (5) for the j th of I subjects. χ^2 is asymptotically chi square with I degrees of freedom (Mood and Graybill, 1963, p.227).

Hence, t tests whether the mean of a single subject's sequence differs from some hypothesized mean, and χ^2 tests the deviations across subjects. To test for the deviations being consistent in sign the statistic

$$T = \sum_{j=1}^I (t_j) / \sqrt{I} \quad (34)$$

is standard normal when N is large (Feller, 1971, p.46). To compare two sequences with no carry-over the statistic is

$$t = [(X1.-X2.) - (u1-u2)] / \sqrt{(\text{Var}1/N1 + \text{Var}2/N2)} , \quad (35)$$

where the parameters are obtained from each sequence separately.

Unfortunately, the choice of K can sometimes be crucial. The basic rule is to choose K small compared to N but large enough to include the $c(k)$ which are large with respect to $c(0)$. K cannot be large or Var becomes inconsistent and when $K=N-1$ then $\text{Var}=0$ (Norman, 1971). Furthermore, since $c(k)$ can be positive or negative an inappropriate choice can bias the tests in either direction. Dusoir (1974, 1980) constructed conservative tests by choosing K separately for each subject in the range $0 \leq K \leq 20$ so that Var was a maximum: a useful procedure since the results rejected the null hypothesis.

Procedures identical to Dusoir's (1974, 1980) were used to test the probability matching hypothesis over the last

100 trials of both sessions in the present experiment. The chi square values for individual subjects appear in Table 4 and 5 for the continuous and intermittent groups, along with $P(R1)$, $P(R1/L,S)$, and d' values. The results are rather striking. Only 6 of the 30 subjects do not show reliable departures from matching over the final 100 trials of session 1. For the second session 11 of the 15 intermittent feedback subjects show significant chi squares for deviations from matching compared to only 4 continuous feedback subjects. Nevertheless, these 4 subjects contribute chi squares large enough to reject matching for the three continuous groups combined (chi square = 71.0537, $p < .001$). Of these continuous groups however, only the chi square for $q = .2$, a group which contained 3 of the 4 non-matching subjects, attained significance.

Table 6 presents the individual subject t values and the group T statistics comparing the first and second sessions. The T s show that subjects in each of the continuous groups produced response sequences in session 2 that differed from those in session 1, but owing to a combination of positive and negative individual subject t values the sequences for the two sessions did not differ systematically for any intermittent group. Hence, the present procedures confirm the results of the analyses of covariance.

The probability matching results seem to reject Kac's (1962) single parameter model for continuous feedback. But there are problems. The model is rejected for a value of

Table 4

Proportions of 'first-longer' responses, proportions correct responses with the long, short order, estimates of discriminability, lag k, and matching chi squares for the last 100 trials of sessions 1 and 2 in Experiment 1.

Continuous Feedback

session 1						session 2					
	subject	P(R1)	P(c/L,S)	d'	lag k	chi square	P(R1)	P(c/L,S)	d'	lag k	chi square
q=.2	1	.39	.70	1.01	3	14.2775**	.56	.75	.61	20	17.9285**
	2	.35	.45	.32	20	5.5720*	.38	.60	.70	0	13.7522**
	3	.49	.65	.26	9	26.8686**	.45	.80	1.19	1	20.9417**
	4	.51	.75	.80	17	29.1575**	.20	.55	1.34	3	0
	5	.51	.80	1.00	4	27.4198**	.24	.65	1.48	1	.8600
	group	.45	.67	.68		103.2953**	.40	.67	1.06		53.4824**
q=.5	6	.76	.96	1.60	14	28.4713**	.55	.84	1.63	0	1.0101
	7	.87	.90	.29	1	115.0794**	.40	.52	.63	9	2.4510
	8	.60	.96	2.46	1	4.0850*	.60	1.00	X	10	3.0865
	9	.74	.78	.25	12	13.2189**	.44	.44	0	7	.7519
	10	.78	.90	.87	0	45.6881**	.59	.76	.91	14	1.0892
	group	.75	.90	1.09		206.5426**	.52	.71	X		8.3887
q=.8	11	.76	.84	1.11	0	.8772	.88	.93	.60	2	5.9757*
	12	.75	.78	.37	3	1.2195	.76	.75	.15	1	.8600
	13	.56	.60	.50	7	14.6797**	.84	.84	-.06	3	1.1617
	14	.60	.66	.81	5	9.7089**	.74	.80	.84	20	.9573
	15	.67	.70	.39	0	7.6437**	.82	.84	.31	6	.2299
	group	.67	.72	1.06		34.1290**	.81	.83	.37		9.1826

* p < .05

** p < .01

Individual and group chi squares based on 1 and 5 df, respectively

Table 5

Proportions of 'first-longer' responses, proportions correct responses with the long, short order, estimates of discriminability, lag k, and matching chi squares for the last 100 trials of sessions 1 and 2 in Experiment 1.

Intermittent Feedback

		session 1					session 2				
	subject	P(R1)	P(c/L,S)	d'	lag k	chi square	P(R1)	P(c/L,S)	d'	lag k	chi square
q=.2	16	.43	.60	.54	0	21.5833**	.22	.30	.28	1	.2301
	17	.38	.35	-.10	0	13.7522**	.56	.75	.61	20	46.0761**
	18	.57	.70	.43	8	55.6574**	.73	.95	1.19	0	142.5173**
	19	.43	.45	.06	0	21.5833**	.42	.33	.35	19	9.3369**
	20	.54	.85	1.13	0	46.5383**	.40	.80	1.36	8	13.4773**
	group*	.47	.59	.41		159.1145**	.47	.67	.76		211.6377**
q=.5	21	.63	.78	.82	11	3.9543*	.62	.88	1.54	0	6.1121*
	22	.84	.86	.16	16	47.8546**	.76	.82	.40	2	29.6828**
	23	.63	.70	.37	0	7.2502**	.75	.88	.87	1	30.3030**
	24	.57	.70	.67	8	1.9937	.38	.44	.32	4	5.0576*
	25	.44	.52	.41	3	1.3680	.58	.76	.46	3	1.6948
	group	.62	.71	.49		62.4207**	.62	.76	.78		72.8503**
q=.8	26	.84	.93	1.44	1	1.0976	.69	.84	2.26	1	4.9192*
	27	.75	.75	0	0	1.3333	.50	.59	1.26	20	8.4906**
	28	.44	.51	1.07	20	24.1939**	.77	.85	1.17	0	.5082
	29	.51	.61	1.57	13	18.2306**	.68	.75	.42	6	4.2230*
	30	.54	.66	.29	3	9.1807**	.80	.79	-.24	18	0
	group	.54	.69	.87		54.0352**	.69	.76	.97		18.1409**

* p < .05

** p < .01

Individual and group chi squares based on 1 and 5 df, respectively

TABLE 6

Individual t and group T statistics comparing the proportions of 'first-longer' responses over the last 100 trials of session 1 with the proportions of 'first-longer' responses over the last 100 trials of session 2 in Experiment 1.

<u>Continuous Feedback</u>			<u>Intermittent Feedback</u>		
Subject	t	T	Subject	t	T
$q = .2$	1	-1.721	16	3.245**	
	2	-.328	17	-2.504*	
	3	.512	18	-2.404*	
	4	4.026**	19	.134	
	5	3.686**	20	1.896	
group		2.762**	group		.164
$q = .5$	6	3.016**	21	.123	
	7	6.474**	22	1.168	
	8	0	23	-1.811	
	9	3.137**	24	2.609*	
	10	1.986*	25	-1.749	
group		6.535**	group		.152
$q = .8$	11	-2.230*	26	2.396*	
	12	-1.60	27	2.238*	
	13	-3.846**	28	-3.909**	
	14	-1.577	29	-1.898	
	15	-2.386*	30	-2.232*	
group		4.561**	group		-1.559

* $p < .05$

** $p < .01$ two tailed testwise probabilities.

$q=P(L,S)=.2$ but not for the symmetric case where $1-q=P(S,L)=.2$. Since the failure to reject matching for $q=.8$ may be due solely to the conservatism of Dusoir's tests, a more powerful procedure (Norman, 1971) was used to test the matching hypothesis.

Norman constructed an analysis of variance, based on time series techniques, to test probability matching for data from a probability learning experiment conducted by Friedman, Burke, Cole, Millward, and Estes (1964). The mean $P(R1)$ for the 57 subjects in the experiment was .7967, while q was .8, representing excellent group matching. Norman assumed that Var and $c(k)$ were equal for each subject, so that the denominator of the F ratio was

$$\text{Var.} = c.(0) + 2 \sum_{k=1}^K c.(k) \quad (36)$$

where $c.(k)$ was the mean $c(k)$ for all subjects at the same level of k . The K was chosen so that $c.(K)$ differed significantly from zero according to a t test based on the 57 subjects. Since k was not chosen separately for each subject so that Var was a maximum, then Norman's test will usually be more powerful than Dusoir's. The numerator of the F is $J/(I-1) \sum_{i=1}^I (X_{i.}-X_{..})^2$, where i corresponds to the i th of I subjects and J is the number of trials. Hence F has $I-1$ and IJ degrees of freedom for large J . The F tests whether the individual subject means differ from each other. Norman (1971) was able to reject matching for the Friedman,

et al. data but it is interesting to note that the between subject variance resulting in the large F [$F(56, \text{infinity})=12.5, p<.001$] accounts for only 7 percent of the total variance estimated by ω^2 , confirming the power of the test.

Norman's procedures were applied to the data from each of the groups for which matching was not rejected using Dusoir's conservative test. The only variation was that k was chosen so that $c.(k)$ was at least 10 percent of $c.(0)$. The use of t to select $c.(k)$ was reasonable for the Friedman, et al. data since there were 56 degrees of freedom, compared to only 4 for each of the groups in the present experiment. Hence, Norman's selection process seemed too conservative for the present data. Nevertheless, the value of k selected by Norman's procedure included only those $c.(k)$ which were 10 percent of $c.(0)$, so 10 percent was chosen as the criterion for the present data. It should be noted that neither this, nor Norman's criterion, has formal statistical justification.

The individual subject $c(k)$ and group mean $c.(k)$ for $k \leq 10$ are presented in Table 7 for the two groups in question. For $q=.5$, k was 7; for $q=.8$ k was 0 according to the 10 percent criterion. Hence, $F(4, \text{infinity})=2.34$, and 2.21 for the $q=.5$ and $.8$ groups, respectively, again failing to reject matching for these groups. However, it is evident that these F 's are not far from the critical .05 rejection value of $F=2.37$. Hence, a different choice of k could lead

TABLE 7

First ten $c.(k)$ for each subject and the group mean, $c.(k)$, over the last 100 trials of session 2 for continuous feedback groups with $q = .5$ and $q = .8$ in Experiment 1.

		Subjects					$c.(k)$
		6	7	8	9	10	
$q = .5$	Lag k						
	0	.2475	.2400	.2400	.2464	.2419	.2432
	1	-.0345	-.0116	.0024	.0133	.0443	.0028
	2	.0180	.0108	.0048	.0313	.0267	.0183
	3	.0049	-.0108	-.0168	-.0062	.0751	.0092
	4	-.0126	-.0284	-.0204	.0163	.0075	-.0075
	5	.0199	.0100	-.0180	.0243	-.0001	.0072
	6	-.0331	.0184	-.0256	.0224	.0223	.0009
	7	-.0007	.0408	.0628	.0148	.0047	.0245
	8	.0018	.0132	.0052	-.0527	-.0087	-.0082
	9	-.0357	.0416	.0376	-.0146	.0196	.0097
	10	.0122	-.0200	.0100	-.0122	-.0221	-.0064
		Subjects					$c.(k)$
		11	12	13	14	15	
$q = .8$	0	.1056	.1824	.1344	.1924	.1476	.1525
	1	-.0045	.0018	-.0059	.0117	-.0209	-.0036
	2	.0053	-.0288	-.0161	.0236	-.0012	-.0034
	3	-.0136	.0007	.0236	-.0070	-.0116	-.0016
	4	.0062	.0001	-.0150	.0123	.0281	.0063
	5	-.0139	-.0105	-.0137	-.0084	-.0004	-.0094
	6	-.0041	-.0187	-.0023	.0135	.0193	.0015
	7	-.0030	-.0268	-.0010	-.0045	-.0193	-.0110
	8	-.0132	.0150	-.0012	-.0252	.0004	-.0048
	9	.0079	-.0056	.0101	.0341	.0101	-.0064
	10	-.0022	.0038	-.0102	-.0240	-.0102	-.0086

to a different conclusion. In this light, it is important to note that ω^2 for each of these tests was .01, so between subject variance was relatively insignificant.

Relative Effects of Feedback and Drift-back

Estimates of the relative efficiency of feedback were obtained from the equation

$$k = [P(R1) - P_n(R1)]/[q - P(R1)] \quad (37)$$

where k is some fraction of the sum of the no-feedback bias parameters $A1$ and $A2$ in model 2 and the ratio, A_f/A_n in model 3. Mean $P(R1)$ values for each of the intermittent groups over the final 100 trials of sessions 1 and 2 were used as estimators. The estimates assume that Kac's single parameter model is appropriate for feedback trials, which may not be the case, particularly for $q=.2$. Nevertheless, they are of interest given the drift-back hypothesis, so they are presented here. When q was .2, .5, and .8, then k was .015, .033, and .464, respectively, suggesting that the effects of feedback are sometimes quite small compared to bias or TOE. Furthermore, k is relatively large when subjects under-matched in session 1.

This dependence of k on q may reflect the tendency for TOE and feedback to coincide on a greater number of trials when q was .8 compared to the other values of q . To see this, consider a rather simplistic characterization of TOE. Suppose that when TOE is positive, 80 percent of L,S trials,

and only 40 percent of S,L trials are perceived correctly. On other trials, suppose that the pairs are perceived incorrectly, and that these conditional probabilities do not change with q . Over a set of feedback trials then, the perception and feedback agree with respect to which stimulus was longer, on: 1) 72 percent of the trials for $q=.8$; 2) 60 percent for $q=.5$; and 3) only 48 percent when $q=.2$. Hence, when q is .2 the feedback and the perception may disagree more often than they agree, perhaps reducing feedback's effectiveness. In other words, the value of the criterion change parameters in the feedback models may depend on the amount of agreement between feedback and the perception. TOE may then diminish, and in certain circumstances negate the effectiveness of feedback. A diminished feedback effect may also occur in the continuous feedback conditions as well. A large positive TOE when q was .2 may have lessened the effect of feedback, or A_f , to the degree that 400 trials were not enough for the process to reach asymptote. Another consideration is that if the feedback and perception disagree too often, some subjects may not even believe the feedback. Re-examination of Table 4 shows that although matching must be rejected for the continuous group with $q=.2$, two subjects displayed quite respectable asymptotic matching, though they over-matched considerably in session 1. The other three subjects show little if any effect of feedback. Similar differences appear in Table 5 among the intermittent feedback subjects. It appears then, that

although figure 3 shows no effect of the intermittent schedule, this may be due to positive and negative changes in $P(R1)$ that are sometimes statistically reliable according to time series analysis. Hence, intermittent feedback may be effective for some subjects.

Between subject differences like these were reported by Dusoir (1980) who suggested that perhaps different models fit different subjects. A potentially more useful proposition, however, suggests that for some subjects and conditions feedback may be totally ineffective, rather than modelled improperly by a single parameter error-correcting formulation. Hence, it may prove to be more worthwhile to investigate the conditions under which feedback has some effects before attempting to distinguish between models of the feedback process.

An important aspect of this argument is that the absence of an intermittent feedback effect in the present experiment may be due to differences in feedback's effectiveness among subjects, not to drift-back. Experiment 2 examines drift-back, and intermittent and continuous feedback extensively in individual subjects. Drift-back is investigated more directly since feedback is removed for entire sessions.

Chapter IV

Experiment 2

In Experiment 2 all possible transitions among continuous, intermittent, and no feedback conditions were examined extensively in individual subjects. If the drift-back hypothesis is correct, then transitions from continuous feedback to no feedback should result in performance changes opposite to those induced by the introduction of feedback. Experiment 2 also provided an opportunity to examine the effects of intermittent feedback more closely. Against expectation, virtually no effects of the intermittent schedule were observed in Experiment 1. It was hoped that by repeatedly providing an individual subject with continuous, intermittent and no-feedback, some effects of intermittency might be observed. In addition, the probability of the presentation order <L,S> was varied for individual subjects, allowing repeated testing of the matching hypothesis with the same subject.

Since transitions to the no-feedback condition occurred after continuous and after intermittent feedback, Experiment 2 permitted an examination of the persistence provided by the two schedules. Because of an analogy between feedback and reinforcement, it might be expected that intermittent feedback would produce greater resistance to drift-back, in

the same way that intermittent reinforcement provides greater resistance to extinction (Humphreys, 1939, & Skinner, 1938). Feedback and reinforcement, are admittedly only vaguely analogous, as are drift-back and extinction, and intermittency has not resulted in persistence in discrete trials discrimination experiments (Mackintosh, 1974). Nevertheless, a persistence effect would be enormously helpful in transfer experiments like those presented below. In transfer designs, feedback cannot be presented in the novel situation if the effects are to be attributed to transfer. Hence, drift-back might occur on the no-feedback, transfer trials. If persistence were provided by an intermittent schedule, this problem might be overcome.

Method

Subjects

An undergraduate student who was enrolled in an introductory psychology course, and a graduate student in English were the subjects. In addition to an hourly wage, each subject received \$.01 for each correct response and was fined \$.01 for each error. The psychology student also received course credit. As in Experiment 1, knowledge of the amount earned was withheld until the completion of the experiment.

Procedures

The apparatus, stimuli, and the experimental setting were the ones used in Experiment 1. Each subject participated in 33 sessions that were identical to those in the previous experiment with the following changes: 1) no two sessions were completed on the same day, and 2) the first twenty trials in a session used the feedback schedule of the immediately preceding session, except on the one occasion when base rates were recovered after a change in the value of q . Table 8 presents the sequence of conditions used in successive sessions. As can be seen, q took on two values for each subject in separate phases of the experiment; and base rates were established at each level of q before feedback was introduced, and after a change in q . Once feedback began at a particular level of q , each feedback condition occurred once in each of four cycles of three sessions. The third session of each cycle provided no feedback. In addition, the schedule in the session preceding these "drift-back" sessions was counter balanced in an ABBA scheme over the four cycles. The alternative BAAB scheme was used for the cycles of feedback schedules when q changed. A scheduling error by the experimenter resulted in an ABAB pattern of cycles for subject KC when q was .5 as Table 8 shows. The completion of the experiment required 67 days for subject KC, and 91 days for subject KO.

Table 8

Session	Subject K C		Subject K O	
	q	P(F)	q	P(F)
1	.5	0	.2	0
2	.5	0	.2	0
3	.5	0	.2	0
4	.8	0	.5	0
5	.8	0	.5	0
6	.8	0	.5	0
7	.8	.5	.5	.5
8	.8	1.0	.5	1.0
9	.8	0	.5	0
10	.8	1.0	.5	1.0
11	.8	.5	.5	.5
12	.8	0	.5	0
13	.8	1.0	.5	1.0
14	.8	.5	.5	.5
15	.8	0	.5	0
16	.8	.5	.5	.5
17	.8	1.0	.5	1.0
18	.8	0	.5	0
19	.5	0	.2	0
20	.5	0	.2	0
21	.5	0	.2	0
22	.5	1.0	.2	1.0
23	.5	.5	.2	.5
24	.5	0	.2	0
25	.5	.5	.2	.5
26	.5	1.0	.2	1.0
27	.5	0	.2	0
28	.5	1.0	.2	.5
29	.5	.5	.2	1.0
30	.5	0	.2	0
31	.5	.5	.2	1.0
32	.5	1.0	.2	.5
33	.5	0	.2	0

Results and Discussion

The Matching Hypothesis

Time series analysis was used to test the matching hypothesis for each session in the feedback cycles. The procedures were identical to the ones used in Experiment 1 for the individual subject data and the results are presented in Table 9 for subject KC and Table 10 for subject KO. The tables also include the proportions of 'first--longer' responses and the proportions of correct responses with the $\langle L, S \rangle$ order. In the sessions providing feedback for subject KC, matching can be rejected only for the first two: sessions 7 and 8. In the absence of feedback, matching is rejected for two of the four sessions when q was .8, but none of the sessions when q was .5. Nevertheless, the significant chi squares with no feedback do not represent drift-back since $P(R1)$ over-matches q in one instance, and under-matches in the other. Furthermore, there appears to be no effect of feedback in the $q=.8$ condition since $P(R1)$ is no closer to matching after four sessions with continuous feedback than before feedback's introduction. In addition, all of the sessions with $q=.5$ show over-matching, though none attain statistical reliability.

Table 10 shows quite a different pattern for subject KO. Matching is rejected for all but two of the 12 sessions with $q=.2$, confirming the results of Experiment 1. However,

Table 9

Proportions of 'first-longer' responses, proportions of correct responses with the L,S order, and matching chi squares for subject KC at each level of q and each level of $P(F)$ over the final 100 trials in each session of Experiment 2.

$q = .8$					$q = .5$				
Session	$P(F)$	$P(R1)$	$P(c/L,S)$	Matching chi square	Session	$P(F)$	$P(R1)$	$P(c/L,S)$	Matching chi square
4	0	.78	.83	X	1	0	.71	.86	X
5	0	.73	.81	X	2	0	.58	.78	X
6	0	.85	.93	X	3	0	.63	.84	X
7	.5	.68	.83	5.5999*	19	0	.54	.84	X
8	1.0	.69	.83	5.6369*	20	0	.61	.88	X
9	0	.74	.89	1.8711	21	0	.61	.96	X
10	1.0	.81	.96	.0593	22	1.0	.59	.98	2.3071
11	.5	.74	.89	1.4556	23	.5	.55	.98	.8139
12	0	.71	.86	3.9340*	24	0	.58	.98	1.7512
13	1.0	.72	.86	1.8588	25	.5	.55	.94	1.0101
14	.5	.80	.94	0	26	1.0	.59	.94	2.8689
15	0	.87	.99	3.8426*	27	0	.60	.96	3.1566
16	.5	.77	.93	.2394	28	1.0	.58	1.00	2.3216
17	1.0	.77	.93	.4181	29	.5	.59	1.00	2.7880
18	0	.78	.95	.2331	30	0	.57	.94	1.9992
					31	.5	.58	.96	2.3230
					32	1.0	.55	.96	.9154
					33	0	.56	.96	1.4611

Table 10

Proportions of 'first-longer' responses, proportions of correct responses with the L,S order, and matching chi squares for subject K0 at each level of q and each level of P(F) over the final 100 trials in each session of Experiment 2.

q = .2					q = .5				
Session	P(F)	P(R1)	P(c/L,S)	Matching chi square	Session	P(F)	P(R1)	P(c/L,S)	Matching chi square
1	0	.56	.85	X	4	0	.76	.82	X
2	0	.73	.75	X	5	0	.81	.90	X
3	0	.75	.85	X	6	0	.83	.86	X
19	0	.43	.60	X	7	.5	.75	.82	26.7380**
20	0	.34	.45	X	8	1.0	.61	.80	4.8273*
21	0	.46	.80	X	9	0	.66	.72	9.9333**
22	1.0	.44	.75	10.6769**	10	1.0	.47	.56	.2231
23	.5	.50	.70	14.6341**	11	.5	.76	.86	37.0620**
24	0	.53	.80	43.7178**	12	0	.76	.74	23.9596**
25	.5	.47	.85	11.8566**	13	1.0	.50	.74	0
26	1.0	.32	.75	3.5378	14	.5	.60	.66	3.1807
27	0	.48	.75	26.3134**	15	0	.56	.60	.3506
28	.5	.39	.55	13.1790**	16	.5	.36	.44	8.5070**
29	1.0	.39	.75	9.8278**	17	1.0	.49	.58	.2947
30	0	.50	.85	36.0000**	18	0	.69	.86	16.8771**
31	1.0	.33	.75	2.9959					
32	.5	.32	.45	5.5148*					
33	0	.38	.75	10.2953**					

matching is rejected only for the first of the continuous feedback sessions when q was .5, and drift-back is quite evident since $P(R1)$ increased whenever a no-feedback session followed continuous feedback.

Tables 11 and 12 present some results that are damaging to the matching hypothesis. The t statistic presented was calculated by the formula

$$t = \sum_{i=1}^4 [P_i(R1) - q] / \sqrt{[\sum_{i=1}^4 (\hat{Var})_i / N_i]} \quad (38)$$

where $P_i(R1)$ and $(Var)_i$ are the proportion of 'first--longer' responses and its associated variance, and N_i is the number of trials for the final 100 trials of each session with the same q and same $P(F)$. Equation 38 is a straightforward extension of Norman's (1971) t for two response sequences to i sequences generated by a single subject with no carry-over. As Table 12 shows, subject KO matched only when q was .5 and feedback was continuous. Curiously however, matching is rejected for subject KC in all cases except when feedback was not provided and q was .8, as can be seen in Table 11.

ROC Analysis

Figures 4-13 present rating-scale receiver operating characteristics (ROCs) based on the subject's confidence responses. A separate ROC is presented for each subject in each of the successive phases of the experiment. For each subject, the first two ROCs were generated from the

TABLE 13

Generalized t statistics comparing proportions of 'first-longer' responses under different probabilities of feedback at $q = .2$ and $q = .5$, and collapsed over q for subject KO in Experiment 2. The final 100 trials of each of the four sessions at a particular q and feedback probability are combined for each comparison of column minus row proportions of 'first-longer' responses.

$q = .2$				$q = .5$				collapsed over q			
<u>P(F)</u>				<u>P(F)</u>				<u>P(F)</u>			
	0	.5	1.0		0	.5	1.0		0	.5	1.0
0	X	-1.23	-2.37**		X	-1.21	-3.27**		X	-1.72*	-4.01**
.5		X	-1.04			X	-2.51**			X	-2.40**

* $p < .05$

** $p < .01$

one tailed testwise probabilities

TABLE 12

Deviations from matching, variances of the deviations, and matching t statistics for subject KO in Experiment 2. Data from the final 100 trials of each of the four sessions at a particular level of q and probability of feedback are combined for each test.

	<u>q = .2</u>			<u>q = .5</u>		
	<u>P(Rl)-q</u>	<u>maximum variance</u>	<u>t</u>	<u>P(Rl)-q</u>	<u>maximum variance</u>	<u>t</u> *
P(F)=0	.2725	.2779	10.338**	.1675	.4451	5.021**
P(F)=.5	.22	.4502	6.558**	.1175	.2402	4.795**
P(F)=1.0	.17	4.4695*	4.962**	.0175	.3971	.5554

* $p < .05$ One tailed testwise probabilities

** $p < .01$

pre-feedback data at each level of q ; the third and fifth, from the feedback cycles for each q ; and the fourth, from the base-rate recovery sessions that followed the first four feedback cycles. Data collected with the different feedback schedules are indicated by different symbols on the ROCs for the feedback cycles. Only the first twenty trials from each session are excluded. The curves cumulate response proportions to the L,S presentation order along the ordinate and the S,L order along the abscissa. The proportions are cumulated over the eight, ordered, response-confidence categories from 'certain - first longer' through 'certain - second longer'. All of the curves display the bowed shape and symmetry around the negative diagonal, typical of rating-scale ROCs found in many of the tasks to which SDT is applied.

Improved discrimination performance over the 33 sessions is evident for both subjects since the curves move progressively away from the major diagonal over the five phases. The improvement for subject KC is spectacularly large and may partially explain the lack of feedback and drift-back effects for this subject, since C_n and the representations of the stimulus pairs may have been changing concurrently. Hence, changes in the representations may have compensated in some way, for changes in the location of C_n . In any case, such changes in discriminability complicate analyses with respect to ALMs since the models assume stable representations.

Figure 4. Confidence rating-based ROC curves for sessions 1, 2, and 3 with no feedback, and with $q=.5$ for subject KC. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 5. Confidence rating-based ROC curves for sessions 4, 5, and 6 with no feedback, and with $q=.8$ for subject KC. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 6. Confidence rating-based ROC curves for sessions 7 through 18. Data is from the feedback cycles with $q=.8$ for subject KC. Circles represent sessions with $P(F)=1.0$, triangles represent $P(F)=.5$, and diamonds represent no-feedback sessions. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 7. Confidence rating-based ROC curves for sessions 19, 20, and 21 with no feedback, and with $q=.5$ for subject KC (second no-feedback phase). The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 8. Confidence rating-based ROC curves for sessions 22 through 33. Data is from the feedback cycles with $q=.5$ for subject KC. Circles represent sessions with $P(F)=1.0$, triangles represent $P(F)=.5$, and diamonds represent no-feedback sessions. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 9. Confidence rating-based ROC curves for sessions 1, 2, and 3 with no feedback, and with $q=.2$ for subject KO. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 10. Confidence rating-based ROC curves for sessions 4, 5, and 6 with no feedback, and with $q=.5$ for subject KO. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 11. Confidence rating-based ROC curves for sessions 7 through 18. Data is from the feedback cycles with $q=.5$ for subject KO. Circles represent sessions with $P(F)=1.0$, triangles represent $P(F)=.5$, and diamonds represent no-feedback sessions. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 12. Confidence rating-based ROC curves for sessions 19, 20, and 21 with no feedback, and with $q=.2$ for subject KO (second no-feedback phase). The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

Figure 13. Confidence rating-based ROC curves for sessions 22 through 33. Data is from the feedback cycles with $q=.2$ for subject K0. Circles represent sessions with $P(F)=1.0$, triangles represent $P(F)=.5$, and diamonds represent no-feedback sessions. The proportions represent decreasing confidence for the 'first-longer' response plotted from the lower left to the upper right (i.e., 'certain first longer' through 'certain first shorter').

In figures 6, 8, 11, and 13 there appears to be some effect of feedback on discriminability. Almost without exception, the points from the continuous feedback sessions are outside the points from the other sessions: an effect that cannot be due to simple practice, since the data for the no-feedback points come from sessions that followed the continuous feedback sessions. Hence, the subjects had more, rather than less practice with the stimuli in the no-feedback sessions.

The effects of the feedback schedules on criteria locations can be seen in figures 14-17. Each point represents $P(R1/L, S)$ as the 'hit' rate plotted against $P(R1/S, L)$ as the 'false alarm' rate. A separate point is included for data collected with each of the feedback schedules, and from each of the base-line phases. Separate figures are provided for each subject at each level of q , and probability matching lines are included as a visual aid. Data from entire sessions are included so criterion locations are not always asymptotic for points from the feedback cycles, nevertheless, the figures demonstrate quite nicely some of the effects already considered. Hence, the major effects for subject KC are on discriminability. However, for subject KO feedback appears to move the criterion towards matching, while removing feedback moves it away.

Figure 14. Proportions of 'first-longer' responses for the two presentation orders, plotted in the unit square to represent criterion locations under various feedback conditions. Proportions of 'first-longer' responses are plotted as 'hits' for the L,S order, as 'false alarms' for the S,L order. The pre-feedback point and a line representing matching (negative slope) are included as visual aids. The data is for subject KC with $q=.8$.

Figure 15. Proportions of 'first-longer' responses for the two presentation orders, plotted in the unit square to represent criterion locations under various feedback conditions. Proportions of 'first-longer' responses are plotted as 'hits' for the L,S order, as 'false alarms' for the S,L order. Pre-feedback points and a line representing matching (negative slope) are included as visual aids. The first point represents sessions 1, 2, and 3. The second point represents sessions 19, 20, and 21. The data is for subject KC with $q=.5$.

Figure 16. Proportions of 'first-longer' responses for the two presentation orders, plotted in the unit square to represent criterion locations under various feedback conditions. Proportions of 'first-longer' responses are plotted as 'hits' for the L,S order, as 'false alarms' for the S,L order. The pre-feedback point and a line representing matching (negative slope) are included as visual aids. The data is for subject KO with $q=.5$.

Figure 17. Proportions of 'first-longer' responses for the two presentation orders, plotted in the unit square to represent criterion locations under various feedback conditions. Proportions of 'first-longer' responses are plotted as 'hits' for the L,S order, as 'false alarms' for the S,L order. Pre-feedback points and a line representing matching (negative slope) are included as visual aids. The first point represents sessions 1, 2, and 3. The second point represents sessions 19, 20, and 21. The data is for subject K0 with $q=.2$.

Drift-back Effects

The drift-back effect evident in subject KO's data is treated more extensively in Table 13 and figures 18, 19 and the upper half of figure 20. Comparisons between different probabilities of feedback were accomplished by computing the statistic

$$t = \frac{\sum_{h=1}^2 \left\{ \sum_{i=1}^4 [Pih(R1) - qi] - \sum_{l=1}^4 [Pil(R1) - qi] \right\}}{\sqrt{\left\{ \sum_{h=1}^2 \left[\sum_{i=1}^4 (\hat{V}ih) + \sum_{l=1}^4 (\hat{V}il) \right] \right\}}} \quad (39)$$

for each q separately, and collapsed over q . Here, h and l refer to higher and lower probabilities of feedback, and therefore contain four values each (i.e. one for each feedback cycle) at each of i levels of q . The Vih and Vil are the four $(Var)i/Ni$ for each sequence of 100 trials at the higher and lower probabilities of feedback at each q . Table 13 shows that, for subject KO, continuous feedback reliably reduced deviations from matching compared to no-feedback at each level of q and for the two levels of q combined. Additionally, for the data collapsed over q , continuous feedback was superior to intermittent feedback, which was in turn, superior to no-feedback. Figure 18, which plots the mean deviations from matching as a function of $P(F)$, presents these results graphically for each q .

It is of interest to know whether the drift-back that is apparent in KO's data occurs between sessions or within sessions since the models presented earlier assume that

TABLE 13

Generalized t statistics comparing proportions of 'first-longer' responses under different probabilities of feedback at $q = .2$ and $q = .5$, and collapsed over q for subject KO in Experiment 2. The final 100 trials of each of the four sessions at a particular q and feedback probability are combined for each comparison of column minus row proportions of 'first-longer' responses.

$q = .2$				$q = .5$				collapsed over q			
<u>P(F)</u>				<u>P(F)</u>				<u>P(F)</u>			
	0	.5	1.0		0	.5	1.0		0	.5	1.0
0	X	-1.23	-2.37**		X	-1.21	-3.27**		X	-1.72*	-4.01**
P(F)											
.5		X	-1.04			X	-2.51**			X	-2.40**

* $p < .05$

** $p < .01$

one tailed testwise probabilities

Figure 18. Deviations from probability matching as a function of the proportion of trials with feedback for subject K0. Points plotted are $P(R1)-q$. Closed symbols indicate sessions with the probability of the L,S order (q) at .2, open symbols indicate sessions with $q=.5$.

drift-back is trial dependent. Figure 19 presents information relevant to this by showing the changes from the last 100 trials of sessions with continuous feedback to the first 100 trials (not counting the 20 warm-up trials) of the next session and finally to the last 100 trials of the latter session. The data were combined over the four session to session transitions (2 with $q=.2$ and 2 with $q=.5$) from continuous to no-feedback, and from continuous to intermittent feedback in figure 18. Both plots show increased deviations from matching from the end of one session through the first 100 trials of the next. Since the drift-back could have occurred before the first 100 trials, a between session effect cannot be ruled out. Nevertheless, both plots also reveal a within session effect since the deviations increase from the first to the last block of 100 trials following continuous feedback. Hence, trial dependent drift-back is not unreasonable.

Persistence Effects

Figures 20 and 21 show the changes for the other transitions between different probabilities of feedback. All of the changes are in the appropriate direction except for those in the upper half of figure 20 which indicate no change for the transitions from intermittent to no-feedback, suggesting a persistence effect for intermittent feedback. A generalized t statistic comparing the no-feedback deviations from matching following continuous with those following intermittent feedback was computed using equation

DEVIATION FROM MATCHING

Figure 19. Deviations from matching for transitions from continuous feedback to the other feedback conditions. The first points indicate data from the last 100 trial block in continuous feedback sessions that were followed by a no feedback session (upper panel), and an intermittent feedback session (lower panel). Second and third points indicate the first and last blocks of 100 trials in those following sessions. Points plotted are $P(R1)-q$ collapsed over $q=.5$ and $q=.2$ for subject K0.

DEVIATION FROM MATCHING

Figure 20. Deviations from matching for transitions from intermittent feedback to the other feedback conditions. The first points indicate data from the last 100 trial block in intermittent feedback sessions that were followed by a continuous feedback session (upper panel), and an no-feedback session (lower panel). Second and third points indicate the first and last blocks of 100 trials in those following sessions. Points plotted are $P(R1)-q$ collapsed over $q=.5$ and $q=.2$ for subject K0.

DEVIATION FROM MATCHING

Figure 21. Deviations from matching for transitions from no-feedback to the other feedback conditions. The first points indicate data from the last 100 trial block in no-feedback sessions that were followed by a continuous feedback session (upper panel), and an intermittent feedback session (lower panel). Second and third points indicate the first and last blocks of 100 trials in those following sessions. Points plotted are $P(R1)-q$ collapsed over $q=.5$ and $q=.2$ for subject K0.

39. Here, $h=\{1,2\}$ and $l=\{1,2\}$ for the final 100 trials of the two no-feedback sessions that followed continuous and intermittent feedback, respectively, and $i=\{1,2\}$ for the two levels of q . The persistence effect was not statistically reliable ($t=-.5811$).

Carry-over Effects

In experiment 1 it was argued that performance may not have been asymptotic after the 400 trials of session 2. In addition, the estimates of feedback's effectiveness suggested that drift-back was more potent than feedback. Data relevant to these issues is presented in figure 22, which plots the mean absolute deviations from matching as a function of the four cycles of feedback schedules. The data are collapsed over the two levels of q for subject KO, and the means are from the final 100 trials in each session. A separate plot is presented for each probability of feedback, and included for reference is the mean of the last 100 trials of the base-line sessions preceding the introduction of feedback. Hence, each point is based on 200 observations, 100 at $q=.5$ and 100 at $q=.2$. Clearly, subject KO's performance with continuous feedback was not asymptotic after the first 400 trial session, but it stabilized thereafter without matching when q was .2. In the no-feedback plot, the deviations from matching decreased progressively, at least over the first three cycles. This apparent carry-over effect suggests that either the drift-back process was not completed in the 400 trials

Figure 22. Mean absolute deviations from matching as a function of the feedback cycle at each probability of feedback for subject KO. Points plotted are the means of the absolute values of $P(R1)-q$. Data is combined over $q=.2$ and $q=.5$.

without feedback, or that the drift-back parameters changed with increased exposure to feedback. In either case, the feedback must have been or become more potent than the drift-back for this subject. The plot for intermittent feedback suggests the same sort of carry-over, again suggesting either non-asymptotic performance or changing parameter values.

q The lack of carry-over was specifically assumed in the various tests of comparisons between two or more response sequences, hence these tests may have been biased. In addition, the bias would have been in the direction of rejecting the null hypotheses since the variance estimates fail to include the positive covariances between sequences. It appears that this bias would have little effect on most of the comparisons. However, the superiority of intermittent over no-feedback that appears in Table 13 may be spurious. The statistic was therefore recomputed using only the final two feedback cycles where carry-over appears minimal. Again intermittent feedback was reliably superior to no-feedback ($t = -2.71$, $p < .01$).

The results of experiments 1 and 2 correspond quite well. Discounting the data of subject KC, whose remarkable improvement in sensitivity made tests of the various models impossible, matching was not rejected when q was .5 or .8, but was clearly rejected when q was .2. Kac's (1962) single parameter model then fails as a general description of feedback effects in discrimination. Furthermore, TOE seems

22
OF / DE

implicated in the failures of the matching hypothesis, since the $q=.8$ condition should be identical to $q=.2$ except for TOE. Hence, the bias towards R_1 , induced by a $q>.5$, is augmented by the positive TOE. The bias away from R_1 when $q<.5$, opposes the same TOE. The argument is aided by consideration of the pre-feedback data in experiments 1 and 2, where aside from the $q=.5$ condition, $P(R_1)$ levels are even more conservative than matching. Similar effects have been noted by Thomas (1975). In the present case, this conservatism was surprising when q was $.8$, given the large positive TOEs which were expected to increase $P(R_1)$ beyond $.8$. Consider then, three separate forces acting on $P(R_1)$: 1) TOE, which serves to increase $P(R_1)$; 2) feedback, which we will assume, for the present, leads toward matching; and 3) conservatism, which forces $P(R_1)$ toward $.5$ in all cases. Note then that when $q=.5$, both conservatism and feedback act toward matching. Only TOE opposes matching. When $q=.8$, feedback and TOE are forces in the direction of increases in $P(R_1)$, and hence, toward matching. Only conservatism opposes. However when $q=.2$, TOE and conservatism combine, so that feedback alone acts to decrease $P(R_1)$, the forces of conservatism and TOE oppose matching. The suggestion is that feedback can sometimes be outmanned by other factors affecting bias. Clearly, such combined forces complicate ALMs, but any theory of feedback effects in discrimination will have to account for the findings that led to the foregoing considerations.

Importantly, the prediction from models 2 and 3 that intermittent feedback should have some effects on $P(R1)$ was demonstrated in experiment 2. Unfortunately, the models, which rely on Kac's matching model for feedback trials, must be rejected simply because matching is rejected for continuous feedback. However, the important and surprisingly large drift-back phenomenon, which was implicated in the ineffectiveness of intermittent feedback in experiment 1, was directly demonstrated in experiment 2, providing a new elaboration of the dynamics of decision processes in psychophysics. Again, TOE is implicated in the attractiveness of the pre-feedback criterion. These drift-back effects and the failures of matching at $q=.2$ demonstrate a real need for a better understanding of TOE and bias effects to which we now turn.

Chapter V

Transfer of Feedback and the TOE

The transfer experiment represents a paradigm common to the learning tradition but seldom used in psychophysics. This is perhaps due to the interests of psychophysicists in the limits of discriminative capacity rather than the relations among psychophysical variables. However, the recent interest and advances in the study of bias and feedback processes suggest that this orientation may be changing and the study of transfer may now attract attention. The logic of the paradigm rests on the adoption of a simple assumption: if the training and test situation have common elements (cf., Thorndike and Woodworth, 1901) then transfer will occur through these elements. Hence, transfer experiments may indicate the degree of similarity between situations. In the following experiments, the transfer of feedback effects is used to examine: 1) the relation of TOE to the difference between the stimuli compared in a length discrimination task, and 2) the relation between TOEs that occur in different types of judgmental tasks -- one that requires the subject to report the direction of the difference between a pair of stimuli (e.g., which is longer?), and one that does not (e.g., same or different?). The first then investigates the similarity in the magnitude of TOE for stimulus pairs of increasing difference by examining

TOE with the larger differences after eliminating TOE for the smaller. The second examines the similarity of the sensory and decisional components of the different tasks by obtaining measures of TOE in one task after eliminating it in the other.

Transfer experiments have a sparse tradition in perception and psychophysics which is discussed by Gibson (1953). Often the experiments examined changes in the Weber fraction for a set of stimuli after practice or training with a different set, sometimes on different continua (e.g., Coover and Angell, 1907). Others, however, dealt specifically with CEs. Juds (1908) for example showed improved accuracy in hitting an underwater target after subjects were informed of the general principle of refraction. Thereafter, the application of general principles became a major interpretation of transfer effects. Hence, Minturn and Reese (1951) found improved estimation of dot patterns after correction training with different patterns. Their subject reported a tendency to commit a CE and an attempt to apply this knowledge throughout the continuum. A model for the correction of a CE was presented by Gibson (1953) which entailed the acquisition of a simple set to shift judgments upward or downward along the continuum. In a sense, ALMs describe in detail the acquisition of such sets through the modification of bias.

The goal of the following experiments is a better understanding of the relation between CEs like the TOE, and bias. A certain dissociation of these factors that complicate discrimination was indicated in a more recent transfer experiment by Jamieson and Petrusic (1976) who found that typically negative TOEs at longer ISIs became positive after TOE was eliminated with feedback at a shorter ISI. The authors argued that TOE decreases in absolute magnitude as ISI increases, and that feedback at the shorter ISI induced a compensating bias that was inappropriately large for the smaller TOE at the longer ISIs. Furthermore, it was suggested that the results support a perceptual interpretation of TOE, as distinct from a criterion bias, since it seemed more reasonable for ISI to affect a perceptual rather than a bias phenomenon. Experiment 3 applies the same reasoning to the difference between the stimuli compared on a discrimination trial. Hence, it is argued that if TOE is a perceptual phenomenon that behaves, for example like contrast, then stimulus difference may well influence the magnitude of the effect. Evidence for an effect of difference was obtained by Cavallaro (Note 1) in a visual length discrimination task, where TOE was smaller at the higher discriminability level. However, if TOE and bias are separable phenomena, as Jamieson and Petrusic (1976) imply, then the effects may have been due more to a failure of isobias than a change in the TOE.

Isobias refers to the constancy of a bias parameter with changes in discriminability. The probability of responding 'no' in a detection task, and the probability of responding 'second' in recognition tasks has been proposed (cf., Dusoir, 1974) as one such parameter, and has been termed $P(R2)$. Hence, in the following sections discussion will switch from the $P(R1)$ of the previous experiments to $P(R2)$. This is done for two reasons: 1) Dusoir (1974) has established a convention for $P(R2)$ in the detection and recognition areas and; 2) in the following experiments TOEs are negative, hence 'second-longer' responses predominate. A number of other bias parameters have been proposed. The most notable of these are the β of classical SDT, $\log b$ of Luce's choice theory, and the more common fixed criterion cutoff C . Parameter C is of particular interest since it has been proposed as an interpretation of CEs in discrimination by Luce and Galanter (1963). In their biased model Luce and Galanter propose a Thurstonian representation for the subjective difference, $X1-X2$, between the first and second stimulus presented on a discrimination trial: the decision rule compares the difference with a criterion C , such that if

$X1-X2 > C$, response $R1$ occurs,

otherwise $R2$,

where R_1 and R_2 are 'first-longer' and 'second-longer' respectively. The bias C , like the difference $X_1 - X_2$, is measured in units of the standard deviation of the difference, hence C relates well to SDT where d' is also based on standard deviation units.

It is worth noting that B , $\log b$, and C each imply that $P(R_2)$ will approach matching as discriminability increases (Dusoir, 1974) so the diminished TOE in Cavallaro's (Note 1) data does not, by itself, indicate a relation between a perceptual TOE and discriminability. To demonstrate the relation, only TOE and discriminability can vary, while other factors like bias are held constant. If the isobias of $P(R_2)$ was assured in situations without TOE (e.g., detection and recognition tasks), the relation between TOE and discriminability might be obtainable. Note that isobias is assumed for $P(R_2) = .5$ by three commonly used bias parameters: 1) the C of Luce and Galanter (1963); 2) β of classical SDT; and 3) $\log b$ of Luce's choice theory.

It was hoped that providing feedback for the smaller difference pairs in experiment 3 would produce matching at $P(R_2) = .5$, and make the isobias assumption more reasonable. This argument is considered further in the following section, along with difficulties that TOE generates for the measurement of subjective length. Most notably, TOEs result in violations of stochastic transitivity. Importantly, if TOEs are constant for all stimulus pairs, the transitivity problem could be easily overcome for a scale of subjective length.

Experiments 4 and 5 concern the same dissociation of TOE and bias. It has been suggested by Blick (1969) that 'same-different' (S-D) and 'longer-shorter' (L-S) judgments of visual extents are mediated by fundamentally different sensory continua. Nevertheless, TOEs occur in both tasks. The TOEs that occur in S-D judgments however, are measured with an index that is presumably bias-free, indicating a sensory rather than bias explanation for TOE. Furthermore, if S-D and L-S tasks are mediated differently, transfer between the tasks should not occur. Hence, the removal of TOE with feedback in one task should have no effect on TOEs in the other. Such a failure of transfer would provide considerable support for a perceptual basis of TOE, in addition to the contention that different sensory continua mediate S-D and L-S judgments.

Chapter VI

Experiment 3

In experiment 3 feedback was provided in a visual length discrimination task for a pair of stimuli differing by a small amount, and TOE assessed for pairs differing by a larger amount. The aim of the experiment was to examine the relation between stimulus difference or discriminability, and TOE, with the effects of bias removed. In addition, it was hoped that the experiment might provide a solution to the serious transitivity problems that result from TOE.

The removal of bias is necessary to the determination of the relation between discriminability and TOE. Consider the data of Cavallaro (Note 1). The TOE, as measured by the index

$$T = .5[P(c/L,S) - P(c/S,L)] \quad (40)$$

decreased with the discriminability of the visual extent pairs. As we noted earlier however, this index can be rewritten as $T = P(R1) - .5$, which is simply the deviation from matching when $q = .5$. Hence, the decrease in the index implied better probability matching at higher levels of discriminability.

However, to conclude that it was TOE that decreased, as an entity distinct from criterion bias, one must assume that

an isobias function held for $P(R1)$. In other words, one must assume that in the absence of TOE, $P(R1)$ is constant across variations in discriminability. Dusoir (1974) argues that evidence for the isobias of $P(R1)$ [or alternatively $P(R2)$ in Dusoir's terms] exists only for the case when $P(R1)=q=.5$ (e.g., Parks, 1966 & Thomas and Legge, 1970), and moreover, Dusoir (1980) presents data against $P(R1)$ for other values of q . With TOE, $P(R1)$ and q differ considerably, even when $q=.5$ as was demonstrated in the first two experiments. Hence, isobias may not hold.

The first two experiments also demonstrated, however, that probability matching can be obtained by providing feedback for judgments typically affected by TOE, hence it is possible to obtain $P(R1)=q=.5$, the very condition for which $P(R1)$ should be constant with variations in discriminability. The feedback for the less discriminable pairs in Experiment 3 provided such conditions, so that any residual TOE for the larger stimulus differences could be attributed to changes in TOE, not to changes in bias.

The importance of the relation between TOE and stimulus difference can be seen by considering the type of transitivity obtainable when TOEs occur. Transitivity requires that, for any relation R and a set of stimuli $\{a, b, \text{ and } c\}$ if aRb , and bRc , then aRc . For the arithmetic relation 'greater than', symbolized $>$, transitivity requires that if $a>b$ and $b>c$, then $a>c$. With fallible data, the relation $a>b$ is equivalent to 'a' being stochastically

dominant over 'b' ($aSDb$). In the context of length discrimination, this means that a is judged longer than b on more than 50 percent of the trials. One minimal condition for the representability of any attribute is called weak stochastic transitivity (WST) which requires that if $aSDb$, and $bSDc$, then $aSDc$. Two other transitivity conditions are defined: 1) moderate stochastic transitivity (MST) which occurs when WST is satisfied, and in addition, the proportion of times 'a' is selected over 'b' [$P(a>c)$ in the present paradigm] exceeds the minimum of $P(a>b)$ and $P(b>c)$; and 2) strong stochastic transitivity (SST) which occurs when WST is satisfied and $P(a>c)$ exceeds the maximum of both $P(a>b)$ and $P(b>c)$.

The difficulties that TOEs cause with respect to the representability of an attribute can be seen in Table 14 which presents data from Cavallaro (Note 1) in which the subjects were required to judge whether the first or second stimulus was the longer (or alternatively, the shorter). The Table presents the proportion of times that the column stimulus was selected over the row stimulus as the longer. The stimuli were three visual extents; 6.72, 7.00, and 7.28 cm in length corresponding to the stimuli X, Y, and Z in Table 14. For subject MW, the upper triangular matrix demonstrates SST, but the lower matrix, where the first stimulus is the longer, fails to satisfy the requirements of WST. Hence, when the presentation order was $\langle L, S \rangle$, the data imply that $X>Y$, and $Y>Z$, but $Z>X$. The intransitivity is

TABLE 14

. Proportions of 'second-longer' responses for the stimulus pairs in the S,L and L,S presentation orders in Cavallaro (Note 1).

P(R2)

		Subject MW					Subject TW		
		Stimulus 2					Stimulus 2		
		X	Y	Z			X	Y	Z
Stimulus 1	X	-	.968	.983	Stimulus 1	X	-	.943	.857
	Y	.658	-	.941		Y	.547	-	.835
	Z	.209	.537	-		Z	.213	.454	-

clearly due to TOE, which appears as an over-estimation of the second stimulus with respect to the first. It is also clear that a scaling routine that subtracted a constant from the second stimulus could generate scale values that were transitive. Furthermore, it appears that subject MW might satisfy SST were it not for the TOE effects, since simply subtracting .16 from each of the proportions would produce SST. The subtraction of a constant would not have the same effect on the upper triangular matrix for subject TW however, since only MST holds. The removal of a progressively decreasing TOE, on the other hand, could satisfy SST.

In experiment 3, subjects discriminated visual extent stimuli for which TOEs are typically negative. Feedback, instead of a scaling routine, was used to eliminate TOE for the embedded pairs of a stimulus triple {X, Y, and Z}. If TOE is constant then the feedback should eliminate the negative TOE for the bounding pair as well, satisfying SST. However, if TOE is inversely related to discriminability, then feedback with the embedded pair should induce a bias that is too large for the bounding pair, and TOE should become positive.

Method

Subjects

The subjects were two female students enrolled in an introductory course in psychology. Again, each received an

hourly wage, and \$.01 for each correct response and was fined \$.01 for each error, and was not informed of her earnings until the completion of the experiment.

Apparatus

The apparatus was the same as in the first two experiments with the following exceptions: 1) the LED was not used 2) the video monitor was raised 10 cm so that the red screen did not obscure the stimuli which were presented in the lower half of the monitor's screen, and 3) a lamp with a 40 watt incandescent bulb, was mounted above and slightly behind the subject in order to eliminate reflections on the video screen.

Stimuli

The stimuli were defined by the horizontal presentation of the end-points of the visual extents on the video monitor. The left and right end-points were 'greater than' (>) and 'less than' (<) symbols respectively, and the position of the left end-point was random over three positions covering a .84 cm range for each stimulus. The visual extents were 6.72, 7.00 and 7.28 cm in length.

Procedures

Subject SG served in 13, and subject PK in 14 sessions of approximately 1.5 hrs each. During the first five sessions all 9 possible pairs from the stimulus triple were presented

without feedback. Hence, there were three identical pairs [(6.72,6.72), (7.00,7.00), and (7.28,7.28)]; four embedded pairs [(6.72,7.00), (7.00,6.72), (7.00,7.28), and (7.28,7.00)] and the two bounding pairs [(6.72,7.28) and (7.28,6.72)]. Only the embedded pairs were presented during sessions 6-8 for subject SG, and during sessions 6-9 for subject PK, with feedback on every trial. All nine pairs were again presented during the subsequent and final five transfer sessions with feedback provided for the embedded pairs only.

One of two phrases formed the instruction for each trial. The subjects were instructed to select the longer stimulus on trials beginning with the phrase 'longer -- first or second?' and the shorter stimulus on trials beginning with the phrase 'shorter -- first or second?'. Each of the 9 stimulus pairs were presented 21 times with each instruction during each of the first five sessions, and 20 times during each of the five transfer sessions. In an attempt to compensate for any possible between session drift-back, an additional four presentations of each embedded pair with feedback began each transfer session. Each of the intervening sessions consisted of 24 presentations of each embedded pair with each instruction.

A typical trial progressed as follows: 1) an instruction was displayed on the video screen for 2.5 sec, and then removed; 2) after a one sec pause the first visual extent stimulus was displayed on the screen and removed one

sec later; 3) following a one sec ISI the second visual extent was displayed and remained on the screen until the subject pressed one of the response buttons; 4) the screen remained blank until the subject again pressed a response button, this time to indicate her confidence in the judgment; 5) on feedback trials the correct response was then displayed on the video screen for 2.5 sec followed by a 2 sec ITI, while on no-feedback trials a 4.5 sec ITI began immediately.

During the first five sessions each instruction by stimulus pair combination appeared once in a random order in the initial 18 trials, and four times in each of five blocks of 72 trials thereafter. The same type of 72 trial blocks was used in the transfer sessions, with the initial 16 trials comprised of two presentations of each of the instruction by stimulus pair combinations using only the embedded pairs. During the intervening sessions each instruction-pair combination appeared three times in the first 24 trials and then 9 times in each of the five blocks of 72 trials. Subjects were allowed a five min rest after completing the initial trials plus one block of 72 trials, and again after completing each of the remaining blocks. No more than two sessions were run in any one day and both subjects completed the experiment in 30 days.

Results and Discussion

Pre-feedback Data Features

Table 15 presents the pre-feedback proportions of 'second - longer' responses $[P(R2)]$ for the two subjects with each stimulus pair in a matrix form. Rows indicate the first stimulus, columns the second stimulus in a pair. Unlike Cavallaro (Note 1) the present subjects satisfy SST in both presentation orders, before feedback, and in spite of TOE. The TOE index $T = .5[P(c/L,S) - P(c/S,L)]$ for the embedded and bounding pairs appears in the upper triangular matrices of Table 16 for each subject. The equivalent index $T = .5 - P(R2)$ appears for the three identical pairs along the diagonals. TOE was negative, large, and diminished as the difference between the stimuli in a pair increased, confirming the findings of Cavallaro (Note 1) and permitting a test of the hypothesis that TOE would become positive for the bounding pair following feedback with the embedded stimulus pairs.

Table 17 presents d' values computed from the data in Table 15. For the embedded and bounding pairs the 'hit' rate was defined as $P(R2/S,L)$ when the presentation order was S,L, and $P(R1/L,S)$ when the order was L,S. The 'false alarm' rate was $P(R2)$ collapsed over the identical pairs for the S,L d' 's, and $P(R1)$ with the identical pairs for the L,S d' 's. As expected the d' 's for the bounding pairs were considerably larger than for the embedded pairs. Furthermore, d' 's for the

TABLE 15

Proportions of 'second-longer' responses for the stimulus pairs in the S,L and L,S presentation orders, and for the identical pairs in the pre-feedback phase of experiment 3.

		<u>P(R2)</u>							
Subject SG					Subject PK				
		<u>Stimulus 2</u>					<u>Stimulus 2</u>		
		6.72	7.00	7.28			6.72	7.00	7.28
Stimulus 1	6.72	.633	.803	.929	Stimulus 1	6.72	.649	.866	.947
	7.00	.421	.643	.805		7.00	.341	.649	.861
	7.28	.143	.346	.633		7.28	.206	.428	.647

TABLE 16

TOE indices for the stimulus pairs in the pre-feedback phase of experiment 3.

$$T = .5 [P(c/L,S) - P(c/S,L)]$$

Subject SGStimulus 2

6.72 7.00 7.28

Stimulus 1

6.72	-.133	-.112	-.036
7.00	X	-.143	-.076
7.28	X	X	-.133

Subject PKStimulus 2

6.72 7.00 7.28

Stimulus 1

6.72	-.149	-.104	-.077
7.00	X	-.149	-.145
7.28	X	X	-.147

TABLE 17

The d' sensitivity indices for the stimulus pairs in the S,L and T,S presentation orders in the pre-feedback phase of experiment 3.

 d'

Subject SGStimulus 2

6.72 7.00 7.28

	6.72	-	.50	1.12
Stimulus 1	7.00	.55	-	.51
	7.28	1.41	.75	-

Subject PKStimulus 2

6.72 7.00 7.28

	6.72	-	.73	1.24
Stimulus 1	7.00	.79	-	.71
	7.28	1.20	.56	-

alternative orders of a pair of stimuli were reasonably symmetrical.

Feedback Effects and Transfer

Figure 23 presents a plot of the mean TOE index for the embedded pairs for each subject during the feedback phase that preceded transfer. The TOE indices for these pairs during the pre-feedback phase are included for comparison purposes. For both subjects TOE was reduced with respect to the pre-feedback phase, and the plots move progressively toward zero TOE and probability matching.

The transfer results appear in Tables 18, 19, and 20, which present for each subject, $P(R2)$, the TOE indices, and d' s respectively, for each stimulus pair collapsed over the final five sessions. The $P(R2)$ data in Table 18 were, not surprisingly, SST in both presentation orders. In addition, overall probability matching was very good since overall $P(R2)$ was .501 for subject SG and .479 for subject PK.

Table 19 demonstrates that the bias induced by feedback with the embedded pairs was not too large for the bounding pairs, since instead of becoming positive, TOE was eliminated for the bounding pairs with both subjects. Only subject PK shows any order effects, but these are small TOEs with one of the embedded pairs and two of the identical pairs. Moreover, these TOEs are positive! Hence, the decreasing TOEs observed before the provision of feedback appear to have resulted more from a failure of the isobias of $P(R2)$ than from a diminishing perceptual effect.

Figure 23. TOE indices during the feedback phase for subject SG (upper panel) and subject PK (lower panel). The index is $P(R2) - .5$.

TABLE 18^a

Proportions of 'second-longer' responses for the stimulus pairs in the S,L and L,S presentation orders, and for the identical pairs in the transfer phase of experiment 3.

<u>P(R2)</u>				
Subject SG				
<u>Stimulus 2</u>				
6.72 7.00 7.28				
Stimulus 1	6.72	.530	.884	.990
	7.00	.116	.465	.873
	7.28	.025	.137	.490
Subject PK				
<u>Stimulus 2</u>				
6.72 7.00 7.28				
Stimulus 1	6.72	.508	.822	.965
	7.00	.097	.442	.803
	7.28	.040	.193	.437

TABLE 19

TOE indices for the stimulus pairs in the transfer phase of experiment 3:

$$T = .5 [P(c/L,S) - P(c/S,L)]$$

Subject SGStimulus 2

6.72 7.00 7.28

	6.72	-0.030	0	-.008
Stimulus 1	7.00	X	+0.035	-.005
	7.28	X	X	+0.010

Subject PKStimulus 2

6.72 7.00 7.28

	6.72	-.008	+0.041	-.003
Stimulus 1	7.00	X	+0.058	+0.002
	7.28	X	X	+0.063

The d' values in Table 20 show reasonably good symmetry for the two presentation orders. A Gourevitch and Galanter (1967) test for the difference between two d' 's was carried out for the two orders of each pair of stimuli, revealing no reliable differences. In addition, note that the sum of the d' 's for the embedded pairs comes close to the d' 's for the bounding pairs in each order. As a crude test of additivity, the difference between the sum of the d' 's for the embedded pairs and the d' of the bounding pair was examined in each presentation order using the Gourevitch and Galanter statistic. None of the differences were reliable ($p < .26$ and $p = 1$ for the L,S and S,L orders respectively for subject SG; and $p < .24$ and $p < .6$ for the L,S and S,L orders for subject PK). Such additivity would imply measurement on an interval scale for d' , and demonstrates the power of the SDT approach to discrimination data.

Relevant for the present purposes however, is the rather clear evidence of the constancy of TOE over various levels of discriminability. Importantly, the Luce and Galanter (1963) treatment of CEs as constant criterion biases provides a reasonable fit to the data. As we have seen, in this conceptualization, bias is a constant C on the scale of subjective differences, and not with respect to $P(R2)$, so that $P(R2)$ converges on matching as discriminability increases (Dusoir, 1974), as occurred in the pre-feedback phase here. As well, an acquired bias that compensated for C would clearly be applied to other stimulus pairs. It is

TABLE 20

The d' sensitivity indices for the stimulus pairs in the S,L and L,S presentation orders in the transfer phase of experiment 3.

		d'						d'			
		<u>Subject SG</u>						<u>Subject PK</u>			
		<u>Stimulus 2</u>						<u>Stimulus 2</u>			
		6.72	7.00	7.28				6.72	7.00	7.28	
Stimulus 1	6.78	-	1.19	2.32		Stimulus 1	6.72	-	.82	1.91	
	7.00	1.19	-	1.13			7.00	1.20	-	.95	
	7.28	1.95	1.08	-			7.28	1.65	.77	-	

worth noting, however, that other bias parameters (e.g., the of classic SDT, and $\log b$ of Luce's choice theory) and models (Broadbent, 1965, 1971; and Smith, 1968) predict similar results for $P(R2)$. In addition, the constant C of Luce and Galanter's model could be reinterpreted as a perceptual effect. Nevertheless, experiment 3 provides no particular support for a perceptual interpretation of time-order effects.

Chapter VII

Transfer of Feedback Between Tasks

The present experiments are concerned with relationships among three common psychophysical discrimination tasks: 1) forced choice (F-C), in which the subject is required to report whether the second of two stimuli was longer or shorter than the first; 2) three-category (3-C), in which the subject may respond longer, shorter, or equal; and 3) same-different (S-D), in which the subject reports only whether the stimuli are the same or different. For the present purposes, F-C and 3-C tasks will be classified together as directional, since they both require the subject to report the direction of any difference between the stimuli in a pair. The specific concern is whether directional and S-D judgments are mediated by a common sensory axis.

The TOEs that appear in F-C tasks are also evident in 3-C (Hellstrom, 1979) and S-D (Blick, 1969) data. In the 3-C task, measurement of the TOE is more difficult since 'same' responses provide no directional information. A number of procedures have been devised. For example, Hellstrom (1979) measures TOE in terms of the width of the 'same' category in standard deviation units, and the Jamieson and Petrusic (1974) TOE index applies if the 'same' responses can be apportioned in some rational manner to the 'longer' and

'shorter' categories. Greenberg's (1965) modification of Thurstone's Law of Comparative Judgment provides one such procedure.

The TOEs that occur in S-D tasks are reflected in a greater portion of the 'same' responses appearing with one rather than the other presentation order. Hence, if TOE is negative, the second stimulus is over-estimated with respect to the first and 'same' responses are more frequent with the L,S than the S,L order. A measure of S-D TOEs with visual extents was used by Blick (1969) who made a strong theoretical assumption about the underlying processes involved in S-D tasks. The index was calculated by first obtaining d's for the two stimulus orders, where 'same' responses were 'hits' when the stimuli were equal, and 'false alarms' when they were unequal, then subtracting the S,L d' from the L,S d'. Blick assumed that these d's were bias-free indices of the 'sameness' of the two visual extents, and cited Wickelgren's (1969) extension of familiarity theory to pitch discrimination as justification. Wickelgren (1969) proposed that the sameness of two tones is given directly, as an immediate sensory quality, and before any information on whether the first was higher or lower than the second. Within familiarity theory then, the sameness in the pitch of two tones is based on the degree of familiarity generated by the second tone with respect to the first. James (1890) expressed a similar position, suggesting that we know that two things are different before

we know how they differ. As evidence for this Wickelgren presents some oddly shaped ROCs from a 3-C (higher-same-lower) pitch discrimination task. The ROCs were non-linear on normal-normal coordinates and suggested bimodal sensory distributions. The two modes presumably reflect an interaction of the two components of the task: first obtaining a sensory value for the sameness of the tones; then, if necessary, determining the direction of any differences. No such evidence has been presented for visual extent however, and Cavallaro (Note 1) obtained 3-C normal-normal ROCs for visual extent that were linear with unit slope. Wickelgren (1969) notes that such plots indicate sensory distributions of differences, rather, than familiarity, that are normal with equal variances.

If familiarity is appropriate for S-D judgments of visual extent however, Blick's (1969) findings represent strong support for the perceptual interpretation ~~of TOEs,~~ since the d' 's she obtained were larger for the L,S pairs than for the S,L pairs. Hence, the L,S order was higher on the familiarity scale than the S,L order, according to a presumably bias-free index.

Another approach to S-D data is due to Hefner (cited in Coombs, 1964; see also Greenberg, 1965, MacMillan, Kaplan, and Creelman, 1977, and Sorkin, 1962; Zinnes and Kurtz, 1968), In this approach S-D, 3-C and F-C are all mediated by the arithmetic difference, $X_1 - X_2$, between the sensory values of the stimuli in the pair. In the unbiased case, the decision rule is:

- 1) if $X_1 - X_2 > C = 0$ respond R_1 , otherwise R_2 , for F-C;
- 2) if $X_1 - X_2 > c_1$ respond R_1 ,
if $X_1 - X_2 < c_2$ respond R_2 , otherwise respond 'same',
where $-c_2 = c_1 = c$, for 3-C;
- 3) if $c_2 < X_1 - X_2 < c_1$ respond 'same', otherwise 'different',
for S-D.

Note that TOEs would occur in these tasks when $C \neq 0$ in F-C, or when $-c_2 \neq c_1$ in S-D and 3-C judgments. TOEs in all three tasks then could be based on biased criteria rather than a perceptual effect.

An important implication of this approach is that feedback effects in directional tasks might transfer to S-D judgments since both are mediated by a common sensory axis. Hence, in terms of ALMs, feedback in 3-C judgments might adjust the criteria c_1 and c_2 used in the decision rule for the S-D task. Furthermore, if we conceive of the interval (c_2, c_1) as an indifference region around the F-C criterion C , then feedback in the F-C task would adjust C and by implication, c_1 and c_2 . Clearly, however, if familiarity mediates S-D, but difference mediates directional judgments, then feedback effects in the directional tasks should not transfer. Two experiments examined whether such transfer takes place.

Chapter VIII

Experiment 4

Experiment 4 examined the transfer of feedback effects between 3-C and S-D judgments. Feedback at the end of a trial informed the subject of the correct response. Hence, 3-C feedback could alert the subject to an error, and in addition, provide information about the direction of the error. According to Kac's error-correcting ALM, such feedback should eliminate TOEs. In addition, if 3-C and S-D tasks are mediated by a common sensory axis, then this feedback effect should transfer to the S-D judgments, eliminating the S-D TOEs. Feedback for S-D judgments however, should not be as effective since feedback on 'different' trials does not inform the subject of the direction of the difference. Without such information, criterion adjustments may occasionally be inappropriate, if they occur at all. S-D feedback then, should eliminate TOEs in neither task.

If S-D judgments are mediated by familiarity, then 3-C feedback should not transfer. A failure of transfer would provide strong support for a perceptual interpretation of TOEs, and in addition underline a basic difference between directional and S-D judgments.

Method

Subjects

The subjects were nine undergraduate students who were enrolled in an introductory psychology course. Each was paid one cent for each correct response and was fined one cent for each error, and in addition, received an hourly wage. No subject was informed of his or her earnings until the completion of the experiment.

Apparatus

The apparatus was the same in all respects to what was used in experiment 3. In addition, the 'yes' and 'no' buttons on the response panel were employed.

Stimuli

Visual extents of 6.72 and 7.00 cm were the stimuli. The extents were again defined by the horizontal display of their endpoints, as in experiment 3.

Procedures

All subjects participated in nine 1.5 hr sessions each consisting of 320 trials divided equally between the 3-C and S-D conditions. Eighty of the 3-C trials began with the instruction 'Same -- or longer -- first or second -- ?', and eighty began with 'Same -- or shorter -- first or second -- ?'. The S-D trials consisted of 80 trials beginning with the instruction 'Same -- yes or no -- ?', and 80 trials beginning with the instruction 'Different -- yes or no -- ?'. Half of the trials in each combination of judgment type by

instruction presented identical pairs, half presented non-identical pairs. The identicals were equally divided between the $\langle 6.72, 6.72 \rangle$, and $\langle 7.00, 7.00 \rangle$ pairs; the non-identicals between the $\langle 6.72, 7.00 \rangle$, and the $\langle 7.00, 6.72 \rangle$ ordered pairs. Each combination of judgment type by instruction by stimulus pair appeared five times in each block of 80 trials in a random order. A five min rest period followed the completion of each 80 trial block. Subjects were instructed to press the 'yes' button if the stimuli were the same on 3-C trials, but if the stimuli were not the same, to select the longer stimulus on 'same or longer' trials, and the shorter on 'same or shorter' trials by pressing one of the buttons marked 'first' or 'second'; and then to press one of the confidence buttons.

A typical trial progressed as follows: 1) an instruction appeared on the video screen for 2.5 sec; 2) after a one sec pause, the first extent was displayed for one sec and then removed; 3) following a one sec ISI, the second extent stimulus was displayed, which remained on the screen until the subject pressed one of the buttons on the response panel; 4) the screen remained blank until the subject pressed a second button on the panel; 5) on feedback trials the correct response was then displayed on the video screen for 2.5 sec, followed by a 2 sec ITI; on no-feedback trials the second response started the timing of a 4.5 sec ITI.

No subject completed more than two sessions in any one day, and the days to completion were a minimum of 14 and a maximum of 29.

Design

The nine subjects were randomly assigned to three groups. No feedback was provided to any of the subjects during the first four sessions. For the final five sessions, one group continued without feedback, one received feedback on 3-C trials, and one received feedback on S-D trials. Hence, there were two feedback groups and a no-feedback control. As well, there was a pre-feedback phase and a feedback phase. Note, however, that the feedback occurred on only half of the trials during the feedback phase. Since experiments 1 and 2 showed that such feedback schedules may require a large number of trials to be effective, it was decided to split the feedback phase into early and late stages. The design then became a 3X3 split-plot factorial with three types of feedback and three feedback stages.

Results and Discussion

Data from the first two sessions were marred by an unacceptable number of inappropriate responses (e.g., 'no' responses to 3-C, and 'first' or 'second responses to S-D instructions) for two subjects. In addition, a failure of the 'yes' and 'no' response buttons resulted in the loss of the major portion of one subject's session 2 data. Sessions 1 and 2 were considered as warm-up therefore, and are not presented. In addition, the fifth session is also omitted since the subjects were becoming familiar with the feedback. The remaining sessions were grouped in pairs for analytical

and graphical purposes. Hence, each data point represents 80 observations with each pair of stimuli in each judgmental task: one just before the introduction of feedback; one just after feedback's introduction; and one from the final feedback sessions.

Difference theory d 's from each stage of the experiment are presented in Table 21 for each subject. The d 's are single point estimates of the distance from the combined identical pairs ($\langle 6.72, 6.72 \rangle$, and $\langle 7.00, 7.00 \rangle$) to each of the non-identical pairs. For the S,L d 's the 'second-longer' response was defined as a 'hit' when the presentation was S,L, and a 'false alarm' when the stimuli in the pair were identical. For the L,S d 's the 'second-longer' response was a 'hit' when the stimuli were identical, and a 'false alarm' when the pair was L,S. On occasion the 'second-longer' response category was empty or contained only one response for the L,S order. When this occurred the 'first-longer' response was used. This procedure proved unsatisfactory on three occasions because the 'first-longer' category was equally deficient for the identicals. These d 's, based on one or two responses, are starred in Table 21. Figure 24 plots the mean d 's combined over presentation order for the three groups over the three feedback stages, and shows little change through the course of the experiment.

TOE indices for the 3-C trials appear in Table 22 for each subject. The index is simply $T = .5 - P(R2)$ computed for

TABLE 21

Difference theory d' 's obtained from the three-category data for each presentation order in each stage of experiment 4.

<u>d'</u>						
<u>NO FEEDBACK</u>						
Subject	<u>Sessions 3 & 4</u>		<u>Sessions 6 & 7</u>		<u>Sessions 8 & 9</u>	
	L,S	S,L	L,S	S,L	L,S	S,L
1	.67	.48	.64	.81	.78	.83
2	1.22*	1.31	1.64	1.51	1.47	1.75
3	1.07	1.25	1.56	1.25	1.19	1.28
<u>FEEDBACK ON S-D</u>						
Subject	<u>pre-feedback Sessions 3 & 4</u>		<u>early feedback Sessions 6 & 7</u>		<u>late feedback Sessions 8 & 9</u>	
	L,S	S,L	L,S	S,L	L,S	S,L
4	.83	1.20	.70	.57	.78	.75
5	.86	1.10	1.08	1.03	.74	1.16
6	1.01	1.66	1.51	1.38	1.07	2.42
<u>FEEDBACK ON 3-CATEGORY</u>						
Subject	<u>pre-feedback Sessions 3 & 4</u>		<u>early feedback Sessions 6 & 7</u>		<u>late feedback Sessions 8 & 9</u>	
	L,S	S,L	L,S	S,L	L,S	S,L
7	1.61	1.43	1.22	1.31	1.38	1.39
8	1.66	1.39	2.23*	2.03	1.83*	1.37
9	1.21	1.19	.93	1.15	1.01	1.07

* Estimate based on 1 response in 80 trials, or less than or equal to 2 responses in 160 trials.

Figure 24. Mean difference theory d' s over the three feedback stages for the three groups. Data is from 3-C trials only. The d' s were calculated for each order of presentation and then combined over the two presentation orders. For L,S d' s a 'first longer' response was defined as a 'hit' when the order was L,S, and a 'false alarm' when the stimuli were equal. 'Second longer' responses were defined as 'hits' and 'false alarms' for the presentations of the S,L and equal pairs for the S,L d' s. Circles, triangles, and diamonds indicate the groups that received 3-C feedback, S-D feedback, and no feedback respectively.

the identicals and non-identicals separately, with half of the 'same' responses counted as R2s. This procedure was used because empty response categories, especially the 'first-longer' category with the S,L pair, prevented the use of Greenberg's (1965) more sophisticated technique for apportioning the 'same' responses. As Table 22 shows, all TOEs were negative throughout the experiment, showing that the feedback was not entirely effective, or perhaps had not yet produced asymptotic performance for the 3-C feedback group. The effect of feedback on the 3-C TOE can be seen in figure 25 which presents group mean TOE indices for the three groups in each feedback stage. Importantly, only the group that received 3-C feedback reduced TOE. Hence, the directional feedback had an effect, but did not eliminate TOE entirely.

The transfer of the directional feedback effect can be seen in Table 23, which presents familiarity theory d's for each presentation order, and the difference between the L,S and S,L d's (symbolized here as T') as a measure of the S-D TOEs for each subject in each stage of experiment 4. Figure 26 plots the corresponding group means. The negative d's in Table 22 indicate that for a number of subjects the proportion of 'same' responses was higher for the L,S order than for the identicals: an effect due to TOE, not the perversity of the subject. Transfer is indicated by the decrease in the S-D TOE for each of the 3-C feedback subjects from the pre- to late feedback stages. The effect

TABLE 22

Indices of TOP in the three-category task for the three stages of experiment 4. The index is $T = .5 \frac{1}{2} P(R_{same}) + P(R_2)$ for the equal pairs and for the unequal pairs where R_2 is 'second-longer'.

<u>T</u>						
<u>NO FEEDBACK</u>						
Subject	<u>Sessions 3 & 4</u>		<u>Sessions 6 & 7</u>		<u>Sessions 8 & 9</u>	
	equals	unequals	equals	unequals	equals	unequals
1	-.32	-.28	-.31	-.30	-.30	-.28
2	-.18	-.18	-.19	-.20	-.19	-.21
3	-.01	-.02	-.09	-.05	-.12	-.06

<u>FEEDBACK ON S-D</u>						
Subject	<u>pre-feedback Sessions 3 & 4</u>		<u>early feedback Sessions 6 & 7</u>		<u>late feedback Sessions 8 & 9</u>	
	equals	unequals	equals	unequals	equals	unequals
4	-.13	-.15	-.14	-.13	-.09	-.18
5	-.19	-.20	-.22	-.19	-.28	-.26
6	-.05	-.09	-.17	-.14	-.12	-.15

<u>FEEDBACK ON 3-CATEGORY</u>						
Subject	<u>pre-feedback Sessions 3 & 4</u>		<u>early feedback Sessions 6 & 7</u>		<u>late feedback Sessions 8 & 9</u>	
	equals	unequals	equals	unequals	equals	unequals
7	-.12	-.11	-.13	-.09	-.11	-.09
8	-.19	-.18	-.12	-.13	-.11	-.10
9	-.22	-.22	-.16	-.15	-.06	-.06

Figure 25. Mean TOE indices for 3-C trials over the three feedback stages for each of the three groups. The index is $.5 - P(R2)$ for the non-identical pairs, where $P(R2)$ is the proportion of 'second longer' responses. Half of the 'same' responses occurring with these pairs were counted as 'second longer' responses. Circles and triangles indicate the groups that received feedback on 3-C and S-D trials respectively. Squares indicate the no-feedback group.

TABLE 23

Familiarity theory d' 's and TOE indices obtained from the S-D data for each order of presentation in each stage of experiment 4. The TOE index is $T' = d'(L,S) - d'(S,L)$.

<u>T'</u>									
<u>NO FEEDBACK</u>									
<u>Sessions 3 & 4</u>				<u>Sessions 6 & 7</u>			<u>Sessions 8 & 9</u>		
Subject	L,S	S,L	T'	L,S	S,L	T'	L,S	S,L	T'
1	-.40	.56	-.96	-.54	.47	-1.01	-1.03	.88	-1.91
2	-.71	1.26	-1.97	-.38	1.41	-1.79	-.35	1.63	-1.98
3	.46	.59	-.23	.06	.77	-.71	.40	1.46	-1.06

FEEDBACK ON S-D

<u>pre-feedback Sessions 3 & 4</u>				<u>early feedback Sessions 6 & 7</u>			<u>late feedback Sessions 8 & 9</u>		
Subject	L,S	S,L	T'	L,S	S,L	T'	L,S	S,L	T'
4	.14	1.34	-1.20	.53	1.11	-.58	1.12	2.08	-.96
5	.16	.36	-.20	.02	.63	-.61	-.09	.67	-.76
6	-.57	.55	-1.12	-.47	1.03	-1.50	-.46	.77	-1.23

FEEDBACK ON 3-CATEGORY

<u>pre-feedback Sessions 3 & 4</u>				<u>early feedback Sessions 6 & 7</u>			<u>late feedback Sessions 8 & 9</u>		
Subject	L,S	S,L	T'	L,S	S,L	T'	L,S	S,L	T'
7	.65	1.33	-.68	.22	1.25	-1.03	.74	1.09	-.35
8	-.02	1.50	-1.52	-.07	1.39	-1.46	.35	.88	-.53
9	-.59	1.49	-2.08	-.62	.82	-1.44	-.33	1.13	-1.46

Figure 26. Mean familiarity theory TOE indices from the S-D trials over the three feedback stages for the groups. The index is $d'(L,S) - d'(S,L)$ (see text for procedure for calculating d' 's). Circles indicate the 3-C feedback group, triangles indicate the S-D feedback group, and squares indicate the no-feedback group.

appears to result from an improvement in the L,S d's, associated with a corresponding decrease for the S,L order. If feedback affects only the criterion or bias, then these d's are not bias-free. TOEs actually increased for each of the subjects in the no-feedback group, and two of the three S-D feedback subjects. Planned comparisons between the pre- and late stages indicated that the decrease in the S-D TOE measure for the 3-C feedback subjects, and the increase for the no-feedback subjects were statistically reliable [$F(1,12)=6.55$, and 5.58 for the 3-C and no-feedback groups respectively, both $p's < .05$].

The present experiment strongly suggests that directional feedback effects transfer to S-D judgments, and hence that 3-C and S-D judgments share a common sensory axis. In addition, familiarity theory must be rejected for the S-D judgments with the visual extents in the present experiment since it is difficult to see how directional effects could transfer to a non-directional sensory quality. Finally, it appears that S-D TOEs cannot be taken as evidence for purely perceptual TOEs, since these TOEs were not measured with a bias-free index, as Blick (1969) argued.

Although the transfer effects in experiment 4 are reasonably strong, TOE was not eliminated in either task. In addition, transfer between 3-C and S-D may be facilitated in some way by the shared 'same' response category. Experiment 5 used feedback in the F-C (e.g., longer-shorter) task in order to examine the transfer of effects from directional to

S-D judgments. Further, S-D trials were omitted during a portion of experiment 5 with the expectation that feedback in the F-C task alone might eliminate TOE entirely.

Experiment 5

Experiment 5 examined the transfer of feedback effects from F-C to S-D judgments of visual extents. Since, unlike 3-C, the F-C task does not share a common response category with S-D, such transfer would demonstrate more dramatically than experiment 4, a common sensory axis for directional and non-directional judgments.

Method

Subject

A female undergraduate student who was enrolled in an introductory psychology course participated as the only subject. She was paid an hourly rate and received one cent for each correct response and was fined one cent for each error. Knowledge of earnings was withheld until the experiment was completed.

Procedures

The apparatus and stimuli were identical to those in experiment 4. In addition, the trials in experiments 4 and 5 were identically structured in sessions 1-4 and 9-12, except that F-C instructions and feedback were substituted for 3-C instructions and feedback. Hence, on half of the F-C trials

subjects were asked 'longer -- first or second -- ?', and on half 'shorter -- first or second -- ?'. No feedback was provided during the first four sessions. During sessions 9-12 feedback appeared only on F-C trials. During sessions 5-8 each of the four blocks of 80 trials consisted of ten randomly ordered presentations of the four stimulus pairs combined with each of the two F-C instructions. Feedback was presented on each of the trials with L,S and S,L pairs, and took the form of the correct response (e.g., 'first' or 'second') with a repetition of the instruction (e.g., 'longer' or 'shorter'). On trials presenting identical pairs, feedback was simply omitted. Hence, the first four sessions were pre-feedback sessions, sessions 5-8 contained only F-C trials with feedback, and sessions 9-12 examined transfer of the feedback effects. The subject served in no more than one session per day, and completed the experiment in 20 days.

Results and Discussion

Table 24 presents some basic statistics for subject MN in experiment 5. The data for the pre-feedback and transfer phases are collapsed over pairs of sessions so that each data point represents 80 observations with the L,S and S,L pairs and 160 for the combined identical pairs. The columns of Table 24 include the proportions of 'second-longer' responses [P(R2)], and the proportions of 'same' responses

TABLE 24

Proportions of 'second-longer' responses and proportions of 'same' responses for subject MN in Experiment 5.

		P(R2)			P(same)		
		Stimulus Pairs			Stimulus Pairs		
	<u>Sessions</u>	<u>L,S</u>	<u>S,L</u>	<u>Identicals</u>	<u>L,S</u>	<u>S,L</u>	<u>Identicals</u>
Pre-feedback	1 and 2	.250	.888	.547	.278	.205	.354
	3 and 4	.225	.975	.692	.138	.063	.197
	5	.250	.963	.669	X	X	X
	6	.141	.963	.639	X	X	X
Feedback	7	.250	.963	.619	X	X	X
	8	.150	.875	.550	X	X	X
Transfer	9 and 10	.177	.886	.633	.037	.075	.144
	11 and 12	.090	.963	.566	0	.025	.031

7

for the combined identical pairs and the two ordered pairs. The rows present the combination of session 1 with 2, 3 with 4, sessions 5-8 separately, and the combination of session 9 with 10, and 11 with 12. Most notable is a drop in $P(R2)$ for the L,S pair over sessions, and an unfortunate decrease in the overall proportion of 'same' responses. In addition, 'same' responses were more frequent for the L,S compared to S,L pair before the introduction of feedback, but less frequent after its introduction, suggesting considerable transfer. The TOE index $T = .5 - P(R2)$ for the F-C task is plotted in figure 27. The proportion $P(R2)$ was computed for all of the stimulus pairs combined. Intermittent feedback was clearly effective since TOE diminished over sessions 5-8, although again TOE was not entirely eliminated.

Familiarity theory d' 's were computed, as in experiment 4, for each of the presentation orders from the S-D data in each of the four pairs of sessions. These appear in Table 25, and figure 28 presents the S-D TOE measure $T' = d'(L,S) - d'(S,L)$. Note that for sessions 11 and 12, subject MN never responded 'same' when the L,S pair was presented. The values presented in Table 25 for sessions 11 and 12 were obtained by computing the average d' 's for the response 'different' over the four confidence levels. Although the reliability of these final estimates can be questioned, it remains that by far the fewest 'same' responses occurred with the L,S order over all four transfer sessions. Hence, the S-D TOE reversed sign after feedback in the F-C task, again demonstrating transfer.

Figure 27. Forced choice TOE indices for subject MN over the 12 sessions. The index is $T = .5 - P(R2)$ computed for the unequal pairs. Indices from the pre-feedback phase (sessions 1-4) and the transfer phase (sessions 9-12) were computed over pairs of sessions so that each point represents 80 observations with each of the presentation orders.

TABLE 25

Familiarity theory d 's for the two stimulus presentation orders for subject MN in Experiment 5.

	<u>Sessions</u>	<u>L,S</u>	<u>S,L</u>
Pre-feedback	1 and 2	.22	.45
	3 and 4	.24	.68
Transfer	9 and 10	.73	.38
	11 and 12	.54*	.54*

* Since there were no 'same' responses for the L,S order in sessions 11 and 12 the d 's are computed as the mean of d 's calculated at each confidence level for the response 'different'.

Figure 28. Familiarity theory TOE indices computed for each pair of sessions in the pre-feedback phase (sessions 1-4) and the transfer phase (sessions 9-12) for subject MN. Data is from S-D trials only. The index is $d'(L,S) - d'(S,L)$ (see text for details of computing d' 's).

Taken together, experiments 4 and 5 demonstrate that directional feedback can be used by subjects to alter their performance with S-D judgments. This transfer effect suggests that a 'sameness' sensory quality was not used for the S-D judgments in the present experiments. For this reason, familiarity theory may be rejected as a general theory of S-D judgments. Furthermore, the apparent importance of the directional feedback in reducing TOE in F-C, 3-C, and S-D judgments suggests difference theory is a more appropriate model for the three tasks. Of course 'sameness' or familiarity may be given directly and before the directional sensory information, and ignored because the subject must also perform a directional task. Hence, when subjects are asked to make S-D and directional judgments in the same experiment, they may come to base S-D judgments on the signed difference between the first and second stimulus (i.e., $X_1 - X_2$), even though the 'sameness' information is available more directly. Nevertheless, Blick's (1969) data came from subjects who were also required to make directional and S-D judgments. It seems unlikely then, that her subjects were responding only on the basis of 'sameness', so her d's were also inappropriate. Furthermore, if the d's did not reflect a sensory component alone, then the TOE measure did not reflect a purely perceptual TOE.

As in experiment 3, these results do not reject the notion that the TOE has a perceptual basis. They do,

however, reduce the amount of evidence that can be claimed as support for the perceptual notion. In addition, it should be noted that experiment 5 does not represent unqualified support for difference theory. Luce and Galanter (1963) point out that, if normally distributed differences $X_1 - X_2$ are the appropriate sensory continuum for S-D judgments, then S-D data should be completely predictable from F-C and 3-C data. Inexplicably, the S-D TOEs in experiment 5 became positive, while F-C TOEs remained slightly but consistently negative.

Chapter IX

General Discussion

In terms of understanding TOE and its relation to bias the final three experiments make a number of interesting contributions. First, since feedback induced biases in directional judgments transfer to S-D judgments, familiarity is not the proper sensory axis for S-D judgments of visual extent. Secondly, and again because of this transfer, the S-D TOE cannot be taken as evidence for a perceptual rather than bias explanation of order effects. Thirdly, TOE appears to be fairly well characterized as a constant with respect to discriminability. Hence, the reduction in $T=.5-P(R2)$ with increased discriminability is easily explained by failures in the isobias of $P(R2)$ when bias or TOE is present. Finally, the violations of weak stochastic transitivity that occasionally appear as a result of TOEs, can be removed by providing feedback, or hypothetically, by subtracting a constant from the scale value for the second stimulus if TOE is negative.

Within the context of the present experiments the reasons for a distinction between perceptual and bias bases for TOE disappear. One reason for the distinction concerns which variables will affect the phenomenon. It is clear that discriminability does not affect TOE. A second reason concerns the point around which the sensory differences are

distributed. Hence, a perceptual interpretation implies that the differences between the stimuli in a pair are not distributed symmetrically around zero for the two presentation orders. The bias notion implies that they are, but that the subject responds on the basis of a criterion $C \neq 0$. The argument reduces to one about a rational zero point, and whether the subject judges with respect to zero. For the measurement of subjective differences, it doesn't matter. It is worth noting that Wickelgren (1968) places the source of the error before the decision, with a bias applied to each sensory distribution. Hence, Wickelgren's model adopts the position of the perceptual notion with respect to the location of the zero point, but is classified as a bias formulation (Jamieson, 1976) because it is mathematically equivalent to the Luce and Galanter (1963) criterion bias model.

However, if TOE is a bias, it is an important one since:

- (1) as we have seen, it is not conservative;
- (2) it is often large and consistent over subjects (Jamieson and Petrusic, 1974);
- (3) it is not a simple preference for one of the response buttons since it does not reverse when the subject is asked 'which is shorter' rather than 'which is longer' (Jamieson and Petrusic, 1974; and Hellstrom, 1979);
- (4) it changes with the size of the standard, hence TOE is positive for durations less than a second in length, but negative for durations that are seconds long (Jamieson and Petrusic, 1974); and

(5) it appears to vary inversely with ISI (Jamieson and Petrusic, 1976). The ISI result suggests that the perception-bias distinction may retain some relevance for memory processes.

The importance of TOE was also apparent in experiments 1 and 2, where the power of order effects were such that: 1) following feedback's elimination of TOE, the effects returned without feedback; 2) TOE was implicated in the failure of matching when feedback was intermittent; and 3) for one of the conditions of the probability of the L,S presentation order (i.e. $q=.2$), TOE appears to prevent matching when feedback is continuous -- not only for individual subjects, but for the group as a whole. The continuous feedback data then, reject Kac's (1962) matching model in a unique way. Previous tests of ALMs have found excellent matching for the group but not for individual subjects (e.g., Dusoir, 1980). Since only asymptotic results were used, other ALMs were not tested. It is clear, however, that any ALM that applies to discrimination tasks, must account for the problems TOE seems to produce, as must the more cognitive models (Kubovy and Healy, 1977), which have not been considered here. Hence, "a 'lose -- shift, win -- be 'confused' model, with a sprinkling of gambler's fallacy" (Kubovy and Healy, 1980 p.259) must also include "and say 'second-longer' a lot -- unless the durations are short".

Whatever the respective fates of the various ALMs, we have at least observed some of the rich detail of dynamic decision processes in psychophysics. Hence biases, built-up with trial by trial feedback, are systematically dismantled in the absence of feedback. This drift-back effect, as we have called it, was quite potent, and seems largely responsible for the ineffectiveness of intermittent feedback. Why the pre-feedback criterion is so attractive, returns us to the question that began this final discussion, and earlier motivated two separate models of intermittent feedback effects -- is TOE bias or perception? The final three experiments suggested that the bias notion is not unreasonable.

References

- Allan, L.G., and Kristofferson, A.B. Psychophysical theories of duration discrimination. Perception and Psychophysics, 1974, 16, 26-34.
- Atkinson, R.C., Carterette, E.C., and Kinchla, R.A. Sequential phenomena in psychophysical judgments: A theoretical analysis. IRE Transactions on Information Theory, 1962, IT-8, 155-162.
- Atkinson, R. and Kinchla, R. A learning model for forced-choice detection experiments. The British Journal of Mathematical and Statistical Psychology, 1965, 18, 183-206.
- Blick, K.I. Decision and decay processes in the short-term memory for length. Journal of Experimental Psychology, 1969, 82, 224-230.
- Broadbent, D.E. A reformulation of Yerkes-Dodson law. British Journal of Mathematical and Statistical Psychology, 1965, 18, 145-157.
- Bush, R.R., Luce, D., and Rose, R.M. Learning models for psychophysics. In R.C. Atkinson (Ed.), Studies in Mathematical Psychology. Stanford: Stanford University Press, 1964.
- Coombs, C.H. A theory of data. New York: Wiley, 1964.
- Coover, J.F., and Angell, F. General practice effect of special exercise. American Journal of Psychology, 1907, 18, 328-340.

- Dorfman, D.D., and Biderman, M. A learning model for a continuum of sensory states. Journal of Mathematical Psychology, 1971, 8, 264-284.
- Dorfman, D.D., Saslow, C.F., and Simpson, J.C. Learning models for a continuum of sensory states re-examined. Journal of Mathematical Psychology, 1975, 12, 178-211.
- Dusoir, A.E. Thomas and Legge's matching hypothesis for detection and recognition tasks: Two tests. Perception and Psychophysics, 1974, 16, 466-470.
- Dusoir, A.E. Treatments of bias in detection and recognition models: A review. Perception and Psychophysics, 1975, 17, 167-178.
- Dusoir, A.E. Some evidence on additive learning models. Perception and Psychophysics, 1980, 27, 163-175.
- Fechner, G. Elements of psychophysics. (Translated by H.E. Adler). New York: Holt, 1966.
- Feller, W. An Introduction to Probability Theory and its Applications. Vol. 2, New York: Wiley, 1971.
- Friedman, M.P., Burke, C.J., Cole, M., Keller, L., Millward, R.B., and Estes, W.K. Two-choice behavior under extended training with shifting probabilities of reinforcement. In R.C. Atkinson (Ed.), Studies in Mathematical Psychology. Stanford: Stanford University Press, 1964. Pp. 250-316.
- Friedman, M.P., Carterette, E.C., Nakatani, L., and Ahumada, A. Comparisons of some learning models for response bias in signal detection. Perception and Psychophysics, 1968, 3, 5-11.

- Gibson, E.J. Improvement in perceptual judgments as a function of controlled practice or training. Psychological Bulletin, 1953, 50, 401-431.
- Green, D.M., and Swets, J.A. Signal detection theory and psychophysics. New York: Wiley, 1965.
- Greenberg, M.G. A modification of Thurstone's law of comparative judgment to accommodate a judgment category of "equal" or "no difference". Psychological Bulletin, 1965, 64, 108-112.
- Greeno, J.G. Effects of nonreinforced trials in two-choice learning with noncontingent reinforcement. Journal of Experimental Psychology, 1967, 64, 373-379.
- Gourevitch, V., and Galanter, E. A significance test for one parameter isosensitivity functions. Psychometrika, 1967, 32, 25-33.
- Guilford, J.P. Psychometric methods. New York: McGraw-Hill, 1954.
- Hellstrom, A. Time errors and differential sensation weighting. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 460-477.
- Humphreys, L.G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. Journal of Experimental Psychology, 1939, 25, 294-301.
- James, W. The principles of psychology. New York: Holt, 1890.
- Jamieson, D.G. Some feedback-induced effects in psychophysics. Unpublished doctoral dissertation, Carleton University, 1976.

- Jamieson, D.G., and Petrusic, W.M. Presentation order effects in duration discrimination. Perception and Psychophysics, 1975, 17, 197-202.
- Jamieson, D.G., and Petrusic, W.M. Feedback versus an illusion in time. Perception, 1976, 7, 91-96. (a)
- Jamieson, D.G., and Petrusic, W.M. On a bias induced by the provision of feedback in psychophysical experiments. Acta Psychologica, 1976, 40, 199-206. (b)
- Judd, C.H. The relation of special training to general intelligence. Educational Review, 1908, 36, 28-42.
- Kac, M. A note on learning signal detection. IRE Transactions on Information Theory, 1962, IT-8, 126-128.
- Kubovy, M., and Healy, A.F. The decision rule in probabilistic categorization: What it is and how it is learned. Journal of Experimental Psychology: General, 1977, 106, 427-446.
- Kubovy, M., and Healey, A.F. Process models of probabilistic categorization. In T.S. Wallston (Ed.), Cognitive Processes in Choice and Decision Behavior. Hillsdale: Lawrence Erlbaum and Assoc., 1980.
- Lee, W. Decision theory and human behaviour. New York: Wiley, 1971.
- Luce, R.D., and Galanter, E. Discrimination. In R.D. Luce, R.R. Bush, and E. Galanter (Eds.), Handbook of mathematical psychology. Volume I. New York: Wiley, 1963.

- Mackintosh, N.J. The psychology of animal learning. London: Academic Press, 1974.
- Macmillan, N.A., Kaplan, H.L., and Creelman, C.D. The psychophysics of categorical perception. Psychological Review, 1977, 84, 452-471.
- Minturn, A.L., and Reese, T.W. The effect of differential reinforcement on the discrimination of visual number. Journal of Psychology, 1951, 31, 201-231.
- Mood, A.M., and Graybill, F.A. Introduction to the theory of statistics. New York: McGraw-Hill, 1963.
- Norman, M.F. Statistical inference with dependent observations: Extensions of classical procedures. Journal of Mathematical Psychology, 1971, 8, 444-451.
- Norman, M.F. Markov processes and learning models. New York: Academic Press, 1972.
- Norman, M.F. Markovian learning processes. SIAM Review, 1974, 16, 143-162..
- Parks, T.E. Signal-detectability theory of recognition memory performance. Psychological Review, 1966, 73, 44-58.
- Schoeffler, M.S. Theory for psychophysical learning. Journal of the Acoustical Society of America, 1965, 37, 1124-1133.
- Skinner, B.F. The behavior of organisms. New York: Appleton-Century-Crofts, 1938.
- Sklansky, J. Markov chain model of adaptive signal detection. Proceedings of the 1963 Bionics Symposium,

Air Force Systems Command, Wright Patterson, A.F.B.,
Ohio, March, 1963.

Smith, P.T. Cost, discriminability, and response bias.
British Journal of Mathematical and Statistical
Psychology, 1968, 21, 35-60.

Sorkin, R.D. Extension of the theory of signal
detectability to matching procedures in psychoacoustics.
Journal of the Acoustical Society of America, 1962, 34,
1745-1751.

Thomas, E.A.C. On a class of additive learning models:
Error-correcting and probability matching. Journal of
Mathematical Psychology, 1973, 10, 241-264.

Thomas, E.A.C. Criterion adjustment and probability
matching. Perception and Psychophysics, 1975, 18,
158-162.

Thomas, E.A.C., and Legge, D. Probability matching as a
basis for detection and recognition decisions.
Psychological Review, 1970, 77, 65-72.

Thorndike, E.L., and Woodworth, R.S. The influence of
improvement in one mental function upon efficiency of
other functions I. Psychological Review, 1901, 8,
247-261 (a).

Thorndike, E.L. and Woodworth, R.S. The influence of
improvement in one mental function upon efficiency of
other functions II. The estimation of magnitude.
Psychological Review, 1901, 8, 384-395 (b).

Thorndike, E.L. and Woodworth, R.S. The influence of improvement in one mental function upon efficiency of other functions III. Functions involving attention, observation and discrimination. Psychological Review, 1901, 8, 553-564 (c).

Torgerson, W.S. Methods of scaling. New York: Wiley, 1958.

Wickelgren, W.A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. Journal of Mathematical Psychology, 1968, 5, 102-122.

Wickelgren, W.A. Associative strength theory of recognition memory for pitch. Journal of Mathematical Psychology, 1969, 6, 13-61.

Zinnes, J.L., and Kurtz, R. Matching, discrimination, and payoffs. Journal of Mathematical Psychology, 1968, 5, 392-421.

Reference Note

1. Cavallaro, G. Characterizing the Time-Order-Error in Discriminations of Visual Extents. Paper presented at the meetings of the Canadian Psychological Association, Calgary, June 1980.

TABLE A1

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order L,S in each session. Data for group with $P(L,S) = .2$ in each session. Data for group with $P(L,S) = .2$ and continuous feedback in second session.

P(C/L,S)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>	
1	1	.30	.50	.40	.70	.70	.70
	2	.70	.89	.80	.80	.70	.75
	3	.50	.30	.40	.60	.60	.60
	4	.70	.70	.70	.70	.80	.75
	Mean	.55	.60	.58	.70	.70	.70
2	1	.60	.50	.55	.70	.90	.80
	2	.50	.50	.50	.60	.50	.55
	3	.40	.40	.40	.20	.60	.40
	4	.50	.40	.45	.70	.50	.60
	Mean	.50	.45	.48	.55	.63	.59
3	1	.50	.70	.60	.60	1.00	.80
	2	.50	.50	.50	.80	1.00	.90
	3	.60	.70	.65	1.00	.90	.95
	4	.40	.90	.65	.70	.90	.80
	Mean	.50	.70	.60	.78	.95	.86
4	1	.80	.90	.85	.40	.70	.55
	2	.80	.80	.80	.50	.70	.60
	3	.70	.90	.80	.80	.60	.70
	4	.70	.80	.75	.50	.60	.55
	Mean	.75	.85	.80	.55	.65	.60
5	1	.30	.70	.50	.80	.70	.75
	2	.90	.90	.90	.60	.60	.60
	3	.80	.70	.75	.80	.50	.65
	4	.90	.70	.80	.60	.70	.65
	Mean	.73	.75	.74	.70	.63	.66

TABLE A2

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order S,L in each session. Data for group with $P(L,S)=.2$ and continuous feedback in second session.

$P(c/S,L)$

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>			
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>	
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>		
1	1	.79	.53	.65	.60	.53	.56	
	2	.45	.48	.46	.53	.58	.55	
	3	.53	.63	.58	.50	.28	.39	
	4	.70	.68	.69	.58	.40	.49	
	Mean	.61	.58	.60	.55	.45	.50	
2	1	.75	.73	.74	.68	.65	.66	
	2	.65	.70	.68	.63	.70	.66	
	3	.70	.63	.67	.75	.75	.75	
	4	.72	.68	.69	.73	.63	.68	
	Mean	.70	.68	.69	.70	.68	.69	
3	1	.73	.68	.65	.58	.63	.60	
	2	.78	.63	.65	.55	.50	.53	
	3	.83	.60	.71	.43	.65	.53	
	4	.60	.50	.55	.63	.65	.64	
	Mean	.72	.60	.66	.54	.61	.58	
4	1	.45	.40	.43	.63	.56	.59	
	2	.58	.60	.59	.85	.73	.79	
	3	.60	.48	.53	.88	.85	.86	
	4	.55	.55	.55	.90	.88	.89	
	Mean	.54	.51	.53	.81	.75	.78	
5	1	.80	.75	.78	.63	.53	.58	
	2	.78	.58	.68	.78	.80	.79	
	3	.48	.48	.48	.93	.85	.89	
	4	.63	.50	.56	.88	.85	.86	
	Mean	.67	.58	.63	.80	.76	.78	

TABLE A3

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order L,S in each session. Data for group with $P(L,S) = .5$ and continuous feedback in second session.

<u>P(c/L,S)</u>							
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>	
6	1	.92	1.00	.96	.84	.84	.84
	2	1.00	.96	.98	.56	.80	.68
	3	.92	.92	.92	.52	.71	.61
	4	.96	.96	.96	.88	.80	.84
	Mean	.97	.96	.97	.70	.79	.74
7	1	.72	.84	.78	.80	.76	.78
	2	.68	.88	.78	.72	.72	.72
	3	.84	.84	.84	.48	.72	.60
	4	.88	.92	.90	.48	.56	.52
	Mean	.78	.87	.83	.62	.69	.66
8	1	.84	.92	.88	1.00	1.00	1.00
	2	.96	1.00	.98	.96	.96	.96
	3	.96	1.00	.98	1.00	1.00	1.00
	4	.92	1.00	.96	1.00	1.00	1.00
	Mean	.92	.98	.95	.99	.99	.99
9	1	.68	.76	.72	.52	.40	.46
	2	.64	.76	.70	.36	.32	.34
	3	.88	.80	.84	.36	.44	.40
	4	.76	.80	.78	.40	.48	.44
	Mean	.74	.78	.76	.41	.41	.41
10	1	.64	.76	.70	.76	.84	.80
	2	.72	.92	.82	.84	.96	.90
	3	.68	.80	.74	.80	.88	.84
	4	.88	.92	.90	.72	.80	.76
	Mean	.73	.85	.79	.78	.87	.83

TABLE A4

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order S,L in each session. Data for group with $P(L,S) = .5$ and continuous feedback in second session.

<u>P(c/S,L)</u>							
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
6	1	.12	.04	.08	.64	.52	.58
	2	.24	.16	.20	.80	.44	.62
	3	.48	.40	.44	.84	.68	.76
	4	.44	.44	.44	.80	.68	.74
	Mean	.32	.26	.29	.77	.58	.68
7	1	.20	.40	.30	.56	.64	.60
	2	.32	.32	.32	.68	.64	.66
	3	.28	.24	.26	.60	.72	.66
	4	.16	.16	.16	.80	.64	.72
	Mean	.24	.28	.26	.66	.66	.66
8	1	.40	.60	.50	.72	.68	.70
	2	.92	.64	.78	.80	.80	.80
	3	.64	.64	.64	.76	.84	.80
	4	.76	.76	.76	.80	.80	.80
	Mean	.68	.66	.67	.77	.78	.78
9	1	.56	.52	.54	.52	.60	.56
	2	.40	.36	.38	.68	.80	.74
	3	.24	.32	.28	.76	.64	.70
	4	.24	.36	.30	.52	.60	.56
	Mean	.36	.39	.38	.62	.66	.64
10	1	.64	.36	.50	.48	.44	.46
	2	.60	.44	.52	.68	.56	.62
	3	.60	.44	.52	.68	.68	.68
	4	.44	.24	.34	.60	.56	.58
	Mean	.57	.37	.47	.61	.56	.59

TABLE A5

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order L,S in each session. Data for group with $P(L,S) = .8$ and continuous feedback in second session.

<u>P(c/L,S)</u>							
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>	
11	1	.70	.88	.79	.78	.88	.83
	2	.78	.80	.79	.95	.93	.94
	3	.78	.95	.86	.90	1.00	.95
	4	.78	.90	.84	.93	.93	.93
	Mean	.76	.88	.82	.89	.93	.91
12	1	.63	.70	.66	.83	.73	.78
	2	.63	.78	.70	.83	.73	.78
	3	.85	.83	.84	.75	.50	.63
	4	.78	.78	.78	.75	.75	.75
	Mean	.72	.77	.74	.79	.68	.73
13	1	.70	.80	.75	.73	.80	.76
	2	.58	.64	.61	.78	.93	.85
	3	.60	.65	.63	.73	.98	.85
	4	.45	.75	.60	.85	.83	.84
	Mean	.58	.71	.65	.77	.88	.83
14	1	.68	.63	.65	.78	.83	.80
	2	.73	.68	.70	.88	.85	.86
	3	.70	.75	.73	.90	.88	.89
	4	.55	.78	.66	.78	.83	.80
	Mean	.66	.71	.69	.83	.84	.84
15	1	.45	.60	.53	.55	.60	.58
	2	.58	.78	.68	.58	.78	.68
	3	.58	.70	.64	.73	.83	.78
	4	.68	.73	.70	.83	.85	.84
	Mean	.57	.70	.64	.67	.76	.72

TABLE A6

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order, S, P in each session. Data for group with $P(L,S) = .8$ and continuous feedback in second session.

P(c/S,L)

Session 1 - No Feedback

Session 2 - Feedback

Subject	Block	<u>Instruction</u>		Mean			
		Shorter	Longer		Shorter	Longer	Mean
11	1	.60	.70	.65	.60	.40	.50
	2	.50	.40	.45	.50	.50	.50
	3	.50	.60	.55	.40	.30	.35
	4	.40	.70	.55	.40	.20	.30
	Mean	.50	.60	.55	.53	.35	.44
12	1	.70	.60	.65	.20	.50	.35
	2	.60	.40	.50	.30	.40	.35
	3	.40	.30	.35	.20	.40	.30
	4	.20	.50	.35	.10	.30	.20
	Mean	.48	.45	.47	.20	.40	.30
13	1	.60	.10	.35	.50	.30	.40
	2	.70	.50	.60	.30	.30	.30
	3	.90	.30	.60	.20	.30	.25
	4	.70	.50	.60	.10	.20	.15
	Mean	.73	.35	.54	.28	.28	.28
14	1	.80	.90	.85	.70	.40	.55
	2	.40	.50	.45	.50	.50	.50
	3	.60	.30	.45	.60	.50	.55
	4	.60	.70	.65	.40	.60	.50
	Mean	.60	.60	.60	.55	.50	.53
15	1	.90	.40	.65	.70	.60	.65
	2	.90	.70	.80	.40	.70	.55
	3	.30	.30	.30	.40	.30	.35
	4	.50	.40	.45	.30	.20	.25
	Mean	.65	.45	.55	.45	.45	.45

TABLE A7

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order L,S in each session. Data for group with $P(L,S)=.2$ and intermittent feedback in second session.

P(c/L,S)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>	
16	1	.30	.30	.30	.40	.30	.35
	2	.10	.20	.15	.80	.10	.45
	3	.20	.30	.25	.50	.10	.30
	4	.60	.60	.60	.40	.10	.25
	Mean	.30	.35	.33	.53	.20	.36
17	1	.80	.50	.65	.40	.20	.30
	2	.70	.70	.70	.30	.70	.50
	3	.50	.40	.45	.70	.70	.70
	4	.40	.30	.35	.70	.80	.75
	Mean	.60	.48	.54	.53	.60	.59
18	1	.80	.60	.70	.50	.60	.55
	2	.90	.50	.70	.80	.80	.80
	3	.60	.50	.55	.80	.80	.80
	4	.60	.80	.70	1.00	.90	.95
	Mean	.73	.60	.66	.78	.78	.78
19	1	.60	.70	.65	.50	.50	.50
	2	.40	.70	.55	.50	.70	.60
	3	.60	.90	.75	.70	.60	.65
	4	.50	.40	.45	.67	.40	.53
	Mean	.53	.68	.60	.59	.55	.57
20	1	.80	.80	.80	.80	.80	.80
	2	.70	1.00	.85	.80	1.00	.90
	3	.70	.80	.75	.90	.90	.90
	4	.70	1.00	.85	.80	.80	.80
	Mean	.73	.90	.81	.83	.88	.85

TABLE A8

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order S,L in each session. Data for group with $P(L,S)=.2$ and intermittent feedback in second session.

P(c/S,L)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>			
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>	
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>		
16	1	.83	.75	.79	.75	.70	.73	
	2	.80	.80	.80	.73	.75	.74	
	3	.80	.60	.70	.78	.90	.84	
	4	.70	.53	.61	.73	.85	.79	
	Mean	.78	.67	.73	.74	.80	.77	
17	1	.70	.65	.68	.68	.65	.66	
	2	.75	.55	.65	.56	.65	.63	
	3	.65	.73	.69	.43	.48	.45	
	4	.63	.60	.61	.40	.58	.49	
	Mean	.68	.63	.66	.52	.59	.56	
18	1	.50	.53	.51	.53	.40	.46	
	2	.45	.53	.49	.48	.45	.46	
	3	.40	.60	.50	.20	.45	.33	
	4	.45	.48	.46	.33	.33	.33	
	Mean	.45	.53	.49	.38	.41	.40	
19	1	.68	.55	.61	.68	.70	.69	
	2	.68	.53	.60	.78	.63	.70	
	3	.70	.55	.63	.67	.63	.65	
	4	.68	.48	.58	.69	.53	.61	
	Mean	.68	.53	.56	.70	.62	.66	
20	1	.45	.35	.40	.83	.65	.74	
	2	.60	.35	.48	.60	.53	.56	
	3	.53	.58	.55	.58	.68	.63	
	4	.53	.55	.54	.80	.60	.70	
	Mean	.53	.46	.49	.70	.61	.66	

TABLE A9

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order L,S in each session. Data for group with $P(L,S) = .5$ and intermittent feedback in second session.

<u>P(c/L,S)</u>							
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
21	1	.80	.80	.80	.88	.88	.88
	2	.60	.76	.68	.72	.76	.74
	3	.72	.76	.74	.88	.68	.78
	4	.80	.76	.78	.92	.84	.88
	Mean	.73	.77	.75	.85	.79	.82
22	1	.96	.92	.94	.88	.84	.86
	2	.96	.88	.92	.84	.84	.84
	3	.92	.88	.90	.88	.84	.86
	4	.88	.84	.86	.80	.84	.82
	Mean	.93	.88	.91	.85	.84	.85
23	1	.56	.72	.64	.80	.76	.78
	2	.80	.60	.70	.76	.80	.78
	3	.44	.64	.54	.88	.52	.70
	4	.64	.76	.70	.92	.84	.88
	Mean	.61	.68	.65	.84	.73	.79
24	1	.76	.68	.72	.48	.80	.64
	2	.64	.68	.66	.60	.56	.58
	3	.76	.52	.59	.56	.60	.58
	4	.48	.56	.52	.64	.88	.76
	Mean	.66	.61	.64	.57	.71	.64
25	1	.72	.76	.74	.56	.84	.70
	2	.64	.84	.74	.60	.72	.66
	3	.64	.88	.76	.32	.52	.42
	4	.60	.80	.70	.32	.56	.44
	Mean	.65	.82	.74	.45	.66	.56

TABLE A10

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order S,L in each session. Data for group with $P(L,S)=.5$ and intermittent feedback in second session.

<u>P(c/S,L)</u>							
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
21	1	.28	.36	.32	.68	.68	.68
	2	.40	.40	.40	.76	.68	.72
	3	.52	.60	.56	.88	.56	.72
	4	.60	.44	.52	.56	.72	.64
	Mean	.45	.45	.45	.72	.66	.69
22	1	.04	.12	.08	.20	.29	.25
	2	.04	.20	.12	.44	.40	.42
	3	.08	.60	.34	.28	.44	.36
	4	.16	.20	.18	.28	.32	.30
	Mean	.08	.28	.18	.30	.36	.33
23	1	.52	.72	.62	.40	.48	.44
	2	.56	.48	.52	.16	.20	.18
	3	.52	.36	.44	.24	.24	.24
	4	.40	.48	.44	.24	.52	.38
	Mean	.50	.51	.51	.26	.36	.31
24	1	.68	.28	.48	.60	.64	.62
	2	.60	.64	.62	.64	.48	.56
	3	.68	.60	.64	.48	.44	.46
	4	.64	.64	.64	.56	.64	.60
	Mean	.65	.54	.59	.57	.55	.56
25	1	.44	.32	.38	.68	.44	.56
	2	.36	.36	.36	.44	.52	.48
	3	.64	.52	.58	.56	.44	.50
	4	.72	.40	.56	.72	.64	.68
	Mean	.54	.40	.47	.60	.51	.56

TABLE ALL

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order L,S in each session. Data for group with $P(L,S) = .8$ and intermittent feedback in second session.

P(c/L,S)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
26	1	.95	.93	.94	.98	.98	.98
	2	.85	.95	.90	.85	.95	.90
	3	.93	1.00	.96	.93	.83	.88
	4	.88	.98	.93	.80	.88	.84
	Mean	.90	.96	.93	.89	.91	.90
27	1	.80	.88	.84	.63	.80	.71
	2	.58	.70	.64	.65	.75	.70
	3	.83	.68	.75	.73	.80	.76
	4	.70	.80	.75	.60	.58	.59
	Mean	.73	.76	.75	.65	.73	.69
28	1	.58	.58	.58	.68	.85	.76
	2	.75	.58	.66	.85	.73	.89
	3	.48	.65	.56	.85	.90	.88
	4	.50	.53	.51	.78	.93	.85
	Mean	.58	.58	.58	.79	.90	.85
29	1	.55	.48	.51	.70	.78	.74
	2	.53	.63	.58	.73	.68	.70
	3	.63	.68	.65	.73	.85	.79
	4	.53	.70	.61	.70	.80	.75
	Mean	.56	.62	.59	.71	.78	.75
30	1	.88	.78	.83	.65	.80	.73
	2	.65	.70	.68	.65	.78	.71
	3	.70	.65	.68	.85	.75	.80
	4	.65	.68	.66	.85	.73	.79
	Mean	.72	.70	.71	.75	.76	.76

TABLE A12

Each subject's proportions of correct responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions for the presentation order S,L in each session. Data for group with $P(L,S) = .8$ and intermittent feedback in second session.

P(c/S,L)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>	
26	1	.60	.80	.70	.60	.60	.60
	2	.80	.80	.80	.90	.80	.85
	3	.50	.80	.65	.80	1.00	.90
	4	.70	.30	.50	1.00	.80	.90
	Mean	.65	.68	.67	.83	.80	.81
27	1	.60	.60	.60	.50	.50	.50
	2	.70	.70	.70	.60	.40	.50
	3	.60	.50	.65	.90	.70	.80
	4	.20	.30	.25	.80	.90	.85
	Mean	.53	.53	.53	.70	.63	.66
28	1	1.00	.90	.95	.50	1.00	.75
	2	.40	.70	.55	.80	.70	.75
	3	.70	.60	.65	.70	.50	.60
	4	.80	.90	.85	.70	.40	.55
	Mean	.73	.78	.76	.68	.65	.66
29	1	.80	.70	.75	.70	.60	.65
	2	.50	.60	.55	.40	.40	.40
	3	.80	.80	.80	.50	.20	.35
	4	1.00	.80	.90	.70	.50	.60
	Mean	.78	.73	.76	.58	.43	.50
30	1	.50	.30	.40	.20	.70	.45
	2	.40	.50	.45	.30	.40	.35
	3	.50	.40	.45	.20	.10	.15
	4	.50	.40	.45	.20	.10	.15
	Mean	.48	.40	.44	.23	.33	.28

TABLE A13

Each subject's proportions of first-longer responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions. Data for group with $P(L,S)=.2$ and continuous feedback in second session.

P(R1)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>			
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>			<u>Mean</u>
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>		
1	1	.24	.48	.36	.46	.52		.49
	2	.58	.60	.59	.54	.48		.51
	3	.48	.36	.42	.52	.70		.61
	4	.38	.40	.39	.48	.64		.56
2	1	.32	.32	.32	.40	.46		.43
	2	.38	.34	.36	.42	.34		.39
	3	.32	.34	.33	.24	.32		.28
	4	.33	.38	.36	.36	.40		.38
3	1	.32	.40	.36	.46	.50		.48
	2	.28	.40	.34	.52	.60		.56
	3	.26	.46	.36	.66	.46		.58
	4	.40	.58	.49	.44	.46		.45
4	1	.60	.66	.63	.38	.49		.44
	2	.50	.48	.49	.22	.36		.29
	3	.46	.60	.53	.26	.24		.25
	4	.50	.52	.51	.18	.22		.20
5	1	.22	.34	.28	.46	.52		.49
	2	.36	.52	.44	.30	.28		.29
	3	.58	.56	.57	.22	.22		.22
	4	.48	.54	.51	.22	.26		.24

TABLE A14

Each subject's proportions of first-longer responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions. Data for group with $P(L,S)=.5$ and continuous feedback in the second session.

P(R1)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>			
<u>Subject</u>	<u>Block</u>	<u>Instruction</u>		<u>Mean</u>	<u>Instruction</u>		<u>Mean</u>	
		<u>Shorter</u>	<u>Longer</u>		<u>Shorter</u>	<u>Longer</u>		
6	1	.90	.98	.94	.60	.66	.63	
	2	.88	.90	.89	.38	.68	.53	
	3	.72	.76	.74	.34	.51	.43	
	4	.76	.76	.76	.54	.56	.55	
7	1	.76	.72	.74	.62	.56	.59	
	2	.68	.78	.73	.52	.54	.53	
	3	.78	.80	.79	.44	.50	.47	
	4	.86	.88	.87	.34	.46	.40	
8	1	.72	.66	.69	.64	.66	.65	
	2	.52	.68	.60	.58	.58	.58	
	3	.66	.68	.67	.62	.58	.60	
	4	.58	.62	.60	.60	.60	.60	
9	1	.56	.62	.59	.50	.40	.45	
	2	.62	.70	.66	.34	.26	.30	
	3	.82	.74	.78	.30	.40	.35	
	4	.76	.72	.74	.44	.44	.44	
10	1	.50	.70	.60	.64	.70	.67	
	2	.56	.74	.65	.58	.70	.65	
	3	.54	.68	.61	.56	.60	.58	
	4	.72	.84	.78	.56	.62	.59	

TABLE A15

Each subject's proportions of first-longer responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions. Data for group with $P(L,S)=.8$ and continuous feedback in the second session.

<u>P(R1)</u>								
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>			
<u>Instruction</u>					<u>Instruction</u>			
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	
11	1	.64	.76	.70	.70	.82	.76	
	2	.72	.76	.74	.86	.84	.85	
	3	.72	.84	.78	.84	.94	.89	
	4	.74	.78	.76	.86	.90	.88	
12	1	.56	.64	.60	.82	.68	.74	
	2	.58	.74	.66	.80	.70	.75	
	3	.80	.80	.80	.76	.52	.60	
	4	.78	.72	.75	.78	.74	.76	
13	1	.64	.82	.73	.68	.78	.73	
	2	.52	.61	.57	.76	.88	.82	
	3	.50	.66	.58	.74	.92	.83	
	4	.42	.70	.56	.86	.82	.84	
14	1	.58	.52	.55	.68	.78	.73	
	2	.70	.64	.67	.80	.78	.79	
	3	.64	.74	.69	.80	.80	.80	
	4	.52	.68	.60	.74	.74	.74	
15	1	.38	.60	.49	.50	.56	.53	
	2	.48	.68	.58	.58	.68	.63	
	3	.60	.70	.65	.70	.80	.75	
	4	.64	.70	.67	.80	.84	.82	

TABLE A16

Each subject's proportions of first-longer responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions. Data for group with $P(L,S)=.2$ and intermittent feedback in the second session.

<u>P(R1)</u>							
<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
16	1	.20	.26	.23	.28	.30	.29
	2	.18	.20	.19	.38	.22	.30
	3	.20	.38	.29	.28	.10	.19
	4	.36	.50	.43	.30	.14	.22
17	1	.40	.38	.39	.34	.32	.33
	2	.34	.50	.42	.41	.42	.41
	3	.38	.30	.34	.60	.56	.58
	4	.38	.38	.38	.62	.50	.56
18	1	.56	.50	.53	.48	.60	.54
	2	.62	.48	.55	.58	.60	.59
	3	.60	.42	.51	.80	.60	.70
	4	.56	.58	.57	.74	.72	.73
19	1	.38	.50	.44	.36	.34	.35
	2	.34	.52	.43	.28	.44	.36
	3	.36	.54	.45	.41	.42	.41
	4	.36	.50	.43	.38	.46	.42
20	1	.60	.70	.65	.30	.44	.37
	2	.46	.72	.59	.48	.58	.53
	3	.52	.50	.51	.52	.44	.48
	4	.52	.56	.54	.32	.48	.40

TABLE A17

Each subject's proportions of first-longer responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions. Data for group with $P(L,S)=.5$ and intermittent feedback in the second session.

P(R1)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
21	1	.76	.72	.74	.60	.60	.60
	2	.60	.68	.64	.48	.54	.51
	3	.60	.58	.59	.70	.56	.63
	4	.60	.66	.63	.68	.56	.62
22	1	.96	.90	.93	.84	.77	.81
	2	.96	.84	.92	.70	.72	.71
	3	.92	.64	.78	.80	.70	.75
	4	.86	.82	.84	.76	.76	.76
23	1	.52	.50	.51	.70	.64	.67
	2	.62	.56	.59	.80	.80	.80
	3	.46	.64	.55	.82	.64	.73
	4	.62	.64	.63	.84	.66	.75
24	1	.54	.70	.62	.44	.58	.51
	2	.52	.52	.52	.48	.54	.51
	3	.54	.46	.50	.54	.58	.56
	4	.42	.46	.44	.54	.62	.58
25	1	.64	.72	.68	.44	.70	.57
	2	.64	.74	.69	.58	.60	.59
	3	.50	.68	.59	.38	.54	.46
	4	.44	.70	.57	.30	.46	.38

TABLE A18

Each subject's proportions of first-longer responses in blocks of 100 trials for the shorter and longer instructions, and collapsed over instructions. Data for group with $P(L,S)=.8$ and intermittent feedback in the second session.

P(R1)

<u>Session 1 - No Feedback</u>					<u>Session 2 - Feedback</u>		
<u>Instruction</u>					<u>Instruction</u>		
<u>Subject</u>	<u>Block</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>	<u>Shorter</u>	<u>Longer</u>	<u>Mean</u>
26	1	.84	.78	.81	.86	.86	.86
	2	.72	.80	.76	.70	.80	.75
	3	.84	.84	.84	.78	.66	.72
	4	.76	.92	.84	.64	.74	.69
27	1	.72	.78	.75	.60	.74	.67
	2	.52	.62	.57	.60	.72	.66
	3	.74	.64	.69	.60	.70	.65
	4	.72	.78	.75	.52	.48	.50
28	1	.46	.48	.47	.64	.68	.66
	2	.72	.52	.62	.72	.80	.76
	3	.44	.60	.52	.74	.82	.78
	4	.44	.44	.44	.68	.86	.77
29	1	.48	.44	.46	.62	.70	.66
	2	.52	.58	.55	.70	.66	.68
	3	.54	.58	.56	.68	.84	.76
	4	.42	.60	.51	.62	.74	.68
30	1	.80	.76	.78	.68	.70	.69
	2	.64	.66	.65	.66	.74	.70
	3	.66	.64	.65	.84	.78	.81
	4	.62	.66	.64	.84	.76	.80

END

3	0	H	0	8	H	8	2	1
---	---	---	---	---	---	---	---	---

FIN