THE EFFECT OF TYPE OF FEEDBACK ON HUMAN TIMING PERFORMANCE

by

KATHLEEN A. MANGIAPANELLO

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of

the requirements for the degree of Doctor of Philosophy,

The City University of New York

2010

This manuscript has been read and accepted by the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	Chair of Examining Committee Nancy Hemmes
Date	Executive Officer Maureen O'Connor

Nancy Hemmes

Bruce Brown

Lanny Fields

Robert Allan

Brian Rakitin

Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE EFFECT OF TYPE OF FEEDBACK ON HUMAN TIMING PERFORMANCE

by

KATHLEEN A. MANGIAPANELLO

Advisor: Professor Nancy Hemmes

The purpose of the present experiment was to assess the effects of several of forms of feedback on timing performance across a range of durations (range: 5-25 s) under conditions in which participants were required to perform a concurrent task. Type of feedback was characterized in terms of precision, a term adopted to describe the quantitative relation between properties of a given feedback stimulus with properties of the timing response. Two models for describing the precision of feedback stimuli were proposed. In the present experiment, 4 different forms of feedback that varied in level of precision in relation to estimated time and stimulus duration were employed in a between-groups design. Analyses of the log-transformed data suggested that various forms of feedback differentially affected time judgments, indicating a relation between the between precision of feedback and time judgments; greater precision yielded more accurate judgments. One of the two proposed models of feedback precision was better supported by the data. According to that model, precision of feedback increases in direct relation to the number of feedback statements contained within the feedback stimulus.

Acknowledgments

My Family

To my daughter, Kristen Mangiapanello – for all of the hugs and kisses, patience and understanding; for being an amazing little girl who understands things most grown-ups cannot comprehend.

To my husband, Joe Mangiapanello – for making me follow my heart and my dreams. It is because of his love and support that I have reached this goal.

To my parents and my in-laws – for all the help and support over these many years. Thank you for celebrating the good times and helping me through the rough times.

Advisor

Dr. Nancy Hemmes – for all the guidance and support over the many years dedicated in taking me from undergraduate to Ph.D. Many thanks for being a constant source of motivation, and for never giving up on me.

Dissertation Committee

Dr. Bruce Brown – for providing a level of support beyond that which is expected from a committee member. I have learned more from him than I have from any book or in any classroom.

Dr. Lanny Fields – for always being the voice of reason when life seemed out of control. A special thanks for making sure I never forgot how to laugh

Outside Readers

Dr. Robert Allan – for his support of my dissertation project and participation as a member of my defense committee.

Dr. Brian Rakitin – for his support of my dissertation project and participation as a member of my defense committee.

Friends

Pat D'Ateno – for taking this journey with me and for the 24/7 support over these many years. Some of life's most important lessons are to be learned in a white Hyundai.

Moira Cullinan & Liz Goodwin – for making sure Kristen always had a fun and safe place to stay.

Table of Contents

Lists of Tables

Lists of Figures

THE EFFECT OF TYPE OF FEEDBACK ON HUMAN TIMING PERFORMANCE

by

KATHLEEN A. MANGIAPANELLO

In studies of timing with human participants, many factors have been demonstrated empirically to affect the accuracy with which the length of a temporal interval is judged. One variable, however, that has received little attention in the human timing literature is feedback.

One of the earliest studies that investigated the effect of feedback, termed knowledge of results, on human timing was conducted by Robinson (1963). Using a between-groups design, participants were required to produce one of three durations (5, 10 or 15 s) with either no feedback, partial feedback or full feedback. Dependent upon the accuracy of their temporal production, participants in the partial feedback[1] group were presented with one of the following feedback stimuli: "that was too much," "that was not enough" or "that was just right." Those receiving full feedback were told were told the value of their production relative to the temporal stimulus. Robinson found that the magnitude of error in production decreased as a function of feedback at all durations. The performance of participants in the partial and full feedback conditions did not differ; however, both groups were more accurate in their productions than those not receiving feedback. Variability of temporal productions differed for feedback versus no feedback groups; productions became less variable across trials for both feedback groups at all durations, but increased across trials for the no feedback group.

Whereas Robinson (1963) assessed the effects of feedback using a simple production task, Hicks and Miller (1976) tested the effect of feedback on temporal performance by examining the whether the effects of feedback transferred across different ranges of durations. Participants were exposed to two temporal estimation tasks. During the first task, participants were exposed to several blocks of trials consisting of a single range of durations; either short durations (range: 5-10 s) or long durations (range: 40-80 s). Groups of participants were tested either with or without feedback. The second task required participants to judge the other range of durations in the absence of feedback. For those participants receiving feedback during the first task, the sample duration was provided to the participants following each estimate. It was found that mean absolute error in time judgments varied as a function of presence versus absence of feedback. In the second estimation task, participants who were provided with feedback across one range of durations were more accurate in their estimates when they were subsequently exposed to the other range of durations in the absence of feedback, in comparison to groups not provided feedback at all.

Subsequent research by Allen and Clark (1979) examined effects of feedback across a range of durations (5-9 s) similar to the short range used by Hicks and Miller (1976). Allen and Clark required participants to estimate the durations with or without feedback. Participants receiving feedback were verbally provided with the value of the sample duration after each trial. Upon completion of the estimation task, all participants were exposed to a temporal discrimination task without feedback. Participants were required to discriminate between short and long intervals in which a 7-s stimulus served as the standard, and the intervals used in the estimation task served as comparison stimuli. Allen and Clark reported findings for the estimation task that were consistent with the aforementioned studies in that mean absolute error of time estimates was lower under the feedback condition. In addition, performance across trials of the estimation task was less variable for participants receiving feedback than those who did not. Weber fractions showed an increase in discrimination among durations from early to late trials for the feedback group suggesting that sensitivity improved across trials. Conversely, the Weber fraction worsened across trials for the no feedback group. Subsequently, when all participants were further tested without feedback in the temporal discrimination task, there was no significant difference in discrimination measures between the group that had prior exposure to feedback and the group that had not been exposed to feedback, suggesting that the effects of feedback were temporary.

Although there were many procedural differences among the studies cited above; one notable difference is the type of temporal judgment method used—production or estimation. Three experiments conducted by Montare (1988) explored the possibility that the effects of feedback on temporal judgments might covary with the psychophysical method employed. The author reported that regardless of psychophysical method used during testing (i.e., production, reproduction or estimation), participants given feedback were more accurate and less variable in their time judgments than participants not receiving feedback. There were no effects of the psychophysical method per se.

The effects of feedback on human time judgments were also assessed by Kladopoulos, Brown, Hemmes and Cabeza de Vaca (1998), using a modification of the method of reproduction known as the start-stop procedure. Under the start-stop procedure, participants are presented with a sample duration and are required to reproduce the duration by bracketing the sample duration using a two-part response in which a key press is required when the participant judges the duration is about to elapse (start response) and a the release of the key is required when the participant is certain that the duration has elapsed (stop response). The middle time (i.e., midpoint) of the two part response is analogous to the reproduction time in a simple reproduction task.

In their second experiment, Kladopoulos et al. (1998) required different groups of participants to reproduce a 2-s duration with either no feedback, partial (intermittent) feedback (50% of the trials) or full feedback (100% of the trials). The feedback presented to participants contained both verbal and graphical components. The feedback stimulus consisted of all of the following elements presented visually: the numerical target duration, the values of the start and stop responses (to the nearest .01 s) emitted by the participant, and a graph that varied as function of the relation between the sample duration and participant's estimate of the duration, inclusive of the times at which the participant had both pressed and released the response key. The authors found significant differences in middle times in comparing both feedback conditions to the no feedback condition and reported that time judgments approximated sample duration under conditions in which feedback was presented.

Whereas in Experiment 2, feedback effects were assessed for a single duration, in Experiment 4, Kladopoulos et al. (1998) exposed participants to five durations ranging from 2 to 18 s. In this experiment, feedback was presented following 33% of the trials for a single duration in the range; for one group this was the shortest duration in the range (2 s) and for the other, it was the longest duration in the range (18 s). The authors found no differences in mean middle times across the range of durations dependent upon whether feedback was presented for the short or long duration.

In general, in the aforementioned studies, feedback affected timing performance. The research indicates that for the most part, feedback increases accuracy and reduces variability of time judgments. One notable shortcoming of the research described above is that although participants were often instructed not to count the time intervals, no procedures were employed to interfere with chronometric counting. When participants are required to time an interval, they often report counting as a strategy used in judging the value of the interval (see Hemmes, Brown & Kladopoulos, 2004). One technique that has been demonstrated to successfully preclude chronometric counting is the addition of a second task requirement during a timing task. When participants are required to perform a concurrent task, such as a number-reading task, participants report that they are unable to count the value of a temporal interval (see Hemmes et al., 2004).

Evidence of the effects of concurrent task conditions on time judgments was provided in the third of a series of experiments performed by Hemmes et al. (2004). Different groups of participants were exposed to a range of seven durations (2-23 s) under one of three timing tasks: estimation, production or reproduction (start-stop). For all groups, a concurrent number-reading task was presented on 50% of the trials. The authors found that under single-task conditions (i.e., no concurrent task presented), timing was near veridical across the range of durations, regardless of timing method. The relation between sample duration and mean time judgment was linear. In contrast, under dual-task conditions (i.e., concurrent number-reading task was presented), negatively-accelerated power functions were obtained for both the estimation and reproduction groups. Timing was not veridical, and participants tended to underestimate the sample durations, most noticeably at the longest durations. Hemmes et al. argue that it is likely that different processes are involved in timing an interval by counting as opposed to timing an interval when prevented from chronometric counting.

Although feedback effects have been assessed in a number of studies in which timing was the only task required, few studies have explored feedback effects on timing under concurrent task conditions. One such timing study was conducted by Rakitin (2005). In a series of five experiments, participants were presented with a choice time production task that consisted of a time-production task and a concurrent task that was timing related. This task required participants to produce a specified duration by pressing the correct one of four response keys when they judged the duration had elapsed. To do this, they had to select the key that corresponded to the location in which the temporal stimulus had appeared on the screen. In his third experiment, participants were required to produce 3-s intervals using the choice time production task. Rakitin did not provide feedback to one group of participants following choice time productions. Another group received feedback on 25% of the trials. The feedback was presented via computer and consisted of either “too early,” “too late” or “right on time” dependent upon the value of the time judgment. Participants provided significantly longer productions in the absence of feedback compared to when feedback was presented, and the mean productions of the group receiving feedback were very close to veridical across successive blocks of trials.

While the research described above does indeed show that feedback improves the accuracy with which participants judge temporal intervals under a variety of procedures, little consideration is given to the effects of procedural aspects of feedback, including the manner in which feedback is programmed in relation to responding, and parameters of feedback stimuli. This situation is apparent not only in the study of timing behavior, but also in the study of feedback effects in other areas of learning and cognition.

With respect to the behavioral literature, there is no clear definition of feedback stimuli and procedures. According to Peterson (1982), feedback is no more than “professional slang,” yet is often “treated as a principle of behavior.” Ford (1980) refers to feedback as “a cumbersome and disorganized aggregation of methods and procedures” (p.183). With respect to stimuli, both verbal and nonverbal stimuli may serve as feedback (Mallott, 2008). Feedback stimuli have been described as having informational value if they provide *knowledge of results* relative to performance (Kazdin, 1989; Sulzer-Azaroff & Mayer, 1991). Others have proposed that feedback stimuli often identify some criterion level of responding (Prue & Fairbank, 1981). Feedback procedures are often programmed much in the same manner as reinforcement procedures, and feedback stimuli have been likened to reinforcers and punishers (Sulzer-Azaroff & Mayer, 1991) as well as conditioned reinforcers (Kazdin, 1989). The general consensus is that feedback is a response-produced stimulus that is correlated with the emitted response (Catania, 1998; Domjan & Burkhard, 1986).

Given the disparities in the literature in describing feedback at stimulus and procedural level, Mangiapanello and Hemmes (2009) examined feedback from a behavior-analytic perspective. Their review examined feedback applications in a constrained body of literature sampled from the areas of learning, behavior analysis and cognition. Because of the lack of consistency in defining feedback in the studies reviewed, they proposed the following definition of feedback:

> Feedback is a procedure that is operationally defined as *the response-contingent presentation of a stimulus whose properties vary as a function of properties of the antecedent response* (Mangiapanello & Hemmes, 2009).

The authors further state that feedback stimuli can be presented in a number of modalities or forms, and these stimuli may vary with respect to qualitative and/or quantitative properties of responding. Variation in the feedback stimuli may relate to a single response dimension or combination of several dimensions of responding. On the basis of the preceding definition and description of feedback procedures and stimuli, the authors systematically compared feedback procedures with a number of behavioral principles and procedures including, but not limited to, reinforcement (and punishment), shaping, concurrent operants and antecedent procedures. The authors reported that although feedback stimuli and procedures conformed in many ways to these well-established behavioral procedures, there were some notable differences between feedback procedures and typical behavior-analytic procedures; the most noticeable difference pertained to the manner in which feedback stimuli were selected and the procedures were programmed.

Related to the present study, Mangiapanello and Hemmes (2009) discussed differences with respect to the number of values that reinforcers versus feedback stimuli typically assumed along a given continuum. The authors note that in general, reinforcement procedures use a single response criterion for reinforcement, and the values of a reinforcer are dichotomous—present versus absent. Although some feedback manipulations use feedback stimuli that assume dichotomous values, quite often feedback stimuli are programmed to assume more than two values, corresponding to a continuum of response criteria. Accordingly, Mangiapanello and Hemmes (2009) argued that in many instances, feedback varies more *precisely* with properties of responding than does a reinforcer. The authors note that *feedback precision* may be defined in terms of the number of variants of a given feedback stimulus and the manner in which the feedback variants relate to both emitted responding and, in most instances, a predetermined criterion level of responding. Admittedly, the description of precision of feedback provided above is incomplete; there is no theoretical basis for the use of such terminology available in the literature. Rather, the term *precision* is used merely as a tool in an attempt to classify various forms of feedback, potentially leading to a better understanding of how various forms of feedback differentially affect responding.

In terms of the preceding definition of feedback precision, binary (dichotomous) feedback such as “correct” and “incorrect” could be described as low in precision, particularly following incorrect responses. Typically with binary feedback, “incorrect” is associated with a greater number of response variants than is “correct.” Moreover, at times a single response variant is associated with “correct,” whereas all other response variants are associated with “incorrect.” Accordingly, “correct” is more precise than “incorrect.”

Feedback may become more precise as the number of feedback variants programmed increases. For example, as described previously, Robinson (1963) used three feedback variants in her partial feedback condition: “that was just right,” “that was too much” and “that was not enough.” The first was provided following correct judgments of duration, and the latter two for incorrect productions. These stimuli not only describe the accuracy of the response, but classify any errors into two categories – under- versus overproduction of the sample duration. Because there are more feedback variants available, and because the type of errors in responding are differentially classified, as a whole, the type of feedback used by Robinson is more precise than dichotomous feedback.

Many studies have implemented feedback procedures that combine a number of forms of feedback. For example, in a study focusing on estimation of probabilities, Stolarz-Fantino, Fantino and van Borst (2006) programmed five feedback variants that used a combination of feedback corresponding to both the magnitude and direction of errors in responding. When the participants' estimates were within ± 5 points of the correct value of the probability, "within the correct range" was the feedback presented. Estimates that were ± 10 points disparate from the correct probability were followed by feedback stating "slightly above the correct range" or "slightly below the correct range" for estimates that were higher or lower than the correct probability, respectively. When estimated probability deviated from the correct probability by more than 10 points, participants were provided with feedback stating "significantly higher than is correct" or "significantly lower than is correct" (p. 607). The feedback variants used by Stolarz-Fantino et al. described the accuracy of the response, and classified errors in responding into four distinct categories based on both the magnitude and error of the direction of the error. It is argued here that the type of feedback used by Stolarz-Fantino et al. is more precise than the feedback methods previously described.

Several of the timing studies cited earlier in this paper used feedback procedures that provided a yet higher level of precision regarding the discrepancy between emitted responding and criterion responding than was provided by the feedback used by Stolarz-Fantino et al. (2004). For example, in the study conducted by Allen and Clark (1979), upon recording their estimate of an interval, participants were verbally provided with the numeric value of the sample duration. Given the value of the estimate and the value of the sample duration, an exact measure of the discrepancy between emitted and criterion responding is described; that is, when an error occurred, the direction and magnitude of the error could be determined. This feedback is very precise in that there is a one-to-one correspondence between each response variant and each feedback variant.

The level of precision in the feedback stimuli used by Allen and Clark (1979) could be achieved in a number of different ways. For example, the feedback stimuli used in the experiments conducted by Kladopoulos et al. (1998) presented participants with both verbal and graphic feedback. Similar to Allen and Clark, the verbal feedback consisted of the sample duration. The values of the both the start and stop response times were also provided. The graphic component of the feedback stimulus contained no numeric values (i.e., with the exception of "0" to indicate the start of the trial). Rather, the sample duration was marked at the center of a line, and a bracket marked with both "start" and "stop" varied relative to the sample mark; that is, if the start response occurred at a value shorter than the sample duration and the stop response occurred at a value longer than the sample duration, the sample mark fell within the bracket. In contrast, when the stop response was lower in value than the sample duration, the bracket fell to the left of the sample mark. Similarly, if the start response was of greater value than the sample duration, the bracket appeared to the right of the sample mark. The graphic feedback in combination with the start and stop times resulted in a one-to-one relation between each response variant and each feedback variant. One could argue that providing the value of the sample duration was redundant and unnecessary. Nonetheless, the feedback stimuli used by Kladopoulos et al. were much more precise than those used in many of the studies previously described.

As stated throughout this paper, feedback appears in many forms, and presumably plays a role in operant conditioning of a variety of responses, including time estimation. Although many forms of feedback have been reported, no comparisons of the effects of different forms of feedback on performance were located in the feedback literature. There also appears to be an absence of systematic analysis of the logical or formal differences among the various types of feedback. As a result, there is little basis for predicting how different forms of feedback affect performance, particularly time estimation. Based on the foregoing analysis of feedback in terms of precision, the primary goal of the present study was to manipulate feedback in a manner that may be described in terms of a continuum of precision.

In the present study, adult participants were required to produce estimates of temporal stimuli across a range of five durations (5-25 s). In contrast to previous timing studies that incorporated feedback manipulations, participants were required to engage in a concurrent, number-reading task similar to that used by Hemmes et al. (2004) to interfere with chronometric counting. Participants were tested across three experimental sessions, and were provided feedback (if applicable) during the first two sessions, but not during the final session. The effect of feedback was tested between groups, across five feedback conditions. Properties of these feedback stimuli varied with respect to the following properties of the timing response including: the accuracy of the response, the direction of error in responding (i.e., over- vs. underestimation) and the magnitude of error in responding. Each property of the feedback stimuli was tested in isolation as well as in combination with others as an approximate manipulation of precision of the various feedback stimuli. The feedback conditions in the present study included no feedback (no-feedback), binary feedback (binary), magnitude feedback (magnitude), directional feedback (directional) and combined directional and magnitude feedback (combined). Under all conditions (except no-feedback in which the consequence of any estimate produced a "wait" stimulus), a correct estimate yielded the feedback message "correct." Feedback stimuli for incorrect estimates differed across feedback conditions. In the binary condition, the feedback stimulus presented was the message "incorrect." This stimulus correlated only with the accuracy of the response. Following incorrect responding in the magnitude condition, the feedback stimulus presented was a measure of the absolute value of the difference between the sample duration and the estimate (e.g., 3 s). This stimulus indicated that the response was incorrect, and provided a measure of the magnitude of the error in judgment. Under the directional condition, incorrect estimates occasioned feedback stimuli that described the accuracy of the judgment in terms of the direction of the error (over- vs. underestimate) relative to the sample duration. In the final feedback condition, combined, both the direction and the magnitude of the error in judgment were described.

By separating the various properties of the feedback stimuli used in the present experiment, it was possible to construct two indices of the precision of the feedback conditions. The first index described here is based on *statements* under each feedback condition; precision varies directly with the number of feedback statements. For example using a *feedback-statements model*, under the directional feedback condition, an underestimate of sample duration yields the feedback stimulus "Too Short." This stimulus contains two feedback statements in that “Short” describes the direction of the error and “Too” describes accuracy; that is, an error has occurred. Table 1 describes precision of feedback in each condition in terms of the number of feedback statements. The index shown in the table proposes that the precision of a feedback stimulus increases in direct relation to increases in the number of feedback statements presented for each type of feedback. This analysis assumes a given temporal stimulus and an incorrect estimate, and applies to all stimulus/incorrect estimate pairs. It should be noted the incorrect feedback statement in the magnitude, directional and combined conditions is represented by either the words “off by” or “too.” In this analysis, the binary feedback stimuli are characterized as the least precise; the magnitude and directional feedback are equally precise; and combining magnitude and directional feedback results in the highest level of precision. This index predicts that timing performance under all feedback conditions should be more accurate than in the absence of feedback. Temporal estimates should be least accurate in the binary condition, equally accurate in the magnitude and directional conditions, and most accurate in the combined condition.

An alternative index to that presented in the preceding paragraph is described in Table 2. This analysis is based on the number of response-directional changes available on the subsequent trial under each feedback condition, given an incorrect response. Under this *response-directional-change model*, precision varies inversely with the number of changes available. Without feedback, participants may provide an identical estimate, or an increased or decreased estimate of a given sample duration. Assuming that the participant complies with the feedback, there are two possible response alternatives under the binary condition; increase or decrease future estimates. The same holds true under the magnitude condition. In contrast, under the directional and combined conditions, if the participant complies with the feedback, then only one directional response alternative is available for a given sample duration-estimate event. As with the previous index, precision under all feedback conditions is greater than under the no-feedback condition. Feedback under the directional and combined condition is the most precise because the feedback specifies a change in responding in a single direction. Although a change in responding is specified by both the binary and magnitude stimuli, the direction of change is not specified, therefore, these types of feedback are less precise than the directional and combined stimuli. This model predicts that time judgments should be equally accurate under the directional and combined conditions, less accurate under the binary and magnitude conditions and least accurate in the no-feedback condition.

As stated above, under both models, differences in the accuracy of time judgments are expected dependent upon type of feedback presented. It is important to note that it is possible that the effect of type of feedback will depend upon duration. Research has shown that under task conditions, accuracy of time judgments decreases with increases in duration, producing a negatively-accelerated power function (see Hemmes, et al. 2004). Given the preceding, it is likely that differences in timing accuracy related to type of feedback will be most noticeable at the longest durations.

Besides affecting the accuracy of timing responses, it is likely that differences in precision of the various forms of feedback will affect the variability of time judgments. Many of the studies cited in this paper (e.g., Allen & Clark, 1979; Hicks & Miller, 1976; Montare, 1988; & Robinson, 1963), reported decreased variability in time judgments when feedback was presented compared to when feedback was absent. Similarly, in the present study, it is predicted that variability in time judgments will decrease in the presence of feedback. Moreover, if type of feedback affects the accuracy of time judgments it will likely affect variability of temporal estimates; measures of variability should decrease as the precision of feedback increases.

As stated previously, the primary goal of the present study was to perform a feedback manipulation such that a continuum of precision of feedback could be described. A secondary goal of the study was to evaluate whether accuracy of time judgments would differ in comparing performance under conditions in which feedback was presented with subsequent performance following the removal of feedback. Based on the findings of Allen and Clark (1979), it was expected that there would be a decrease in accuracy of time judgments in the final session relative to the first two session during which feedback was presented.

Method

Participants

Fifty Queens College undergraduate students served as participants in fulfillment of an academic requirement for an introductory psychology course. Participants were required to have normal or corrected-to-normal vision, and no color vision anomalies. All participants reported being native-English speakers or learning English by age 6. Participants were required to remove their watches and to turn off any beepers and/or cellular phones in their possession. Participants reporting prior exposure to a timing task were excluded from the present study.

Apparatus and Setting

An IBM Thinkpad® laptop computer with a color monitor was used to present instructions, run experimental sessions and collect data. The timing resolution of the computer has a maximum error of ± 12 μs. The number keys, enter key and space bar served as the manipulanda. A digital microrecorder (Sony®) was used to record participants' vocalizations during all test sessions. Each participant was tested individually in a small office in which all sources of temporal information were removed from the room prior to the onset of testing. The door to the office remained closed during testing.

Design and General Procedure

The present study employed a 5 x 5 x 3 (Feedback x Duration x Session) mixed-factorial design. Feedback was manipulated between groups, whereas session and duration were manipulated within-subjects.

The proposed study was conducted in three sessions across a 90-min time period. Each session lasted approximately 20 min. Training prior to the first session lasted approximately 10 min, and a debriefing session of approximately 10 min followed the final session. There was a 5-min break between sessions. All experimental procedures, including the presentation of instructions (see Appendix A for instructions), were presented by the computer. Each participant was exposed to a single level of the feedback variable (see below) in both Sessions 1 and 2. Session 3 was identical to Sessions 1 and 2 with one exception; feedback was not presented in this session. Participants were exposed to a temporal estimation task during which they were required to engage in a concurrent number-reading task (described below). The number-reading task was presented in all sessions and in all feedback conditions.

Timing Procedures

A temporal estimation task was presented in all sessions. Participants were seated facing the computer monitor with the keyboard located directly in front of the monitor. A trial began with the words, “Press space bar when ready.” One second following this response, a green square measuring 4 cm^2 appeared in the center of the monitor for a specified duration. Once the duration had elapsed, the green square disappeared and was replaced with the words “Your estimate is.” The participants were required to enter their estimate of the duration to the nearest second using the keypad and then press the enter key to register their estimate.

Fifty temporal estimation trials were presented during each session. Participants were exposed to 10 trials of each of five sample durations: 5-, 8-, 11-, 17- and 25-s. per session. The order of sample durations was randomized in 5-trial blocks. All participants experienced the trials in the same order.

During the temporal estimation task, participants engaged in a concurrent, non-temporal task. The task consisted of reading aloud a series of numbers that appeared superimposed on the green rectangle. Participants were required to read aloud these 2- to 4- digit numbers to their full place names; for example, 1124 was read as one thousand, one hundred and twenty-four. Participants were informed that their performance of this task would be recorded. The computer program randomly generated numbers, and the probability of presentation of a 2-, 3- or 4- digit number was equal (i.e., .33). Each number was presented for 400 ms per digit with 65% variability, and the inter-number interval was 300 ms with 65% variability. Similar parameters have been demonstrated to deter participants from counting during temporal tasks (see Hemmes et al., 2004).

Feedback Procedures

Participants were randomly assigned in block random order to one of five feedback conditions (n = 10 participants per condition) that can be characterized as differing in level of precision in relation to the timing response. The five feedback conditions were: no feedback, binary feedback, directional feedback, magnitude feedback and combined directional and magnitude feedback.

No-feedback condition. For each trial, upon entering a time estimate, participants in the no feedback condition were presented with the word “WAIT” on the screen for 4 s. After the 4 s had elapsed, they were prompted to press the space bar to initiate the next trial. The 4-s wait stimulus corresponded to the duration of the feedback stimuli in all other feedback conditions.

Binary feedback condition. Following each estimate by participants in the binary condition, if the estimate equaled the length of the duration, the word "CORRECT" appeared on the screen. All other estimates resulted in the presentation of the word "INCORRECT."

Magnitude feedback condition. In the magnitude condition, if the estimate was equal to the length of the duration, the word "CORRECT" appeared on the screen. All other estimates produced feedback that corresponded to the magnitude of the error in judgment; that is, the absolute value of the difference between their estimate and physical duration. For example, both 2-s and 8-s estimates of a 5-s duration resulted in the feedback message "OFF BY 3 SECONDS."

Directional feedback condition. In the directional condition, the following feedback options were presented. If an estimate was accurate, the word "CORRECT" was presented. If the participant underestimated the duration, the feedback message "TOO SHORT" was presented. Alternatively, if the participant provided an overestimate of the duration, the words "TOO LONG" appeared.

Combined directional plus magnitude feedback condition. Participants in the combined directional plus magnitude condition received feedback that represented both magnitude and direction of error of an estimate (see directional and magnitude feedback above for a full description). Accurate estimates resulted in the presentation of "CORRECT."

Training

Prior to the onset of testing, participants were exposed to eight training trials in which two durations, 3 and 13 s, were each presented on four trials. These two durations were chosen in an attempt to increase the likelihood of participants' entering both correct and incorrect estimates, thereby increasing the probability of exposure to a number of feedback variants, as opposed to a single value of a given form of feedback. The 3-s duration was chosen to increase the likelihood that participants would be exposed to a feedback variant associated with accurate (correct) responding. Research has shown that participants tend to accurately judge durations of less than 5 s, even when the number reading task is present (e.g., see Hemmes et al., 2004). In contrast, the 13-s duration, the arithmetic mean of the durations used in testing, was chosen to increase the likelihood that participants would be exposed to feedback variants associated with inaccurate (incorrect) responding. Participants tend to underestimate values of duration greater than 5 s under concurrent task conditions (e.g., see Hemmes et al., 2004). The training trials were arranged in four blocks of two trials with each duration appearing once in each block in random order. The number-reading task and feedback was presented as described above.

Data Analyses

Estimates greater than 4 *SD* beyond the mean were omitted from the analyses.[2] Eighteen data points were omitted, resulting in 7482 estimates submitted to the initial analysis. Mean time estimates, standard deviation (*SD*) of mean time estimates (see Appendix B, Tables B1-B7) and standard error of mean (*SEM*) time estimates were calculated at each level of each variable and in combination. Upon completion of the analysis of the estimates, log transformations (base 10 log) of these data were performed, and these data were submitted to additional analysis. All testing was conducted at the .05 alpha level. All probability values reported for repeated measures factors (session &

duration) as well as interactions with these factors were corrected using the Huynh-Feldt epsilon.

Results

Analysis of time estimates was accomplished by means of a 5 x 5 x 3 (Feedback x Duration x Session) mixed-factorial ANOVA. The results of this analysis are presented in Appendix B, Table B8. Significant effects of both session, $F(2, 90) = 5.25$, and duration, $F(2, 90) = 463.83$, were found. Contrary to predictions, there was neither a significant effect of feedback, $F(4, 45) = 1.99$, *ns*, nor significant Feedback x Duration interaction, $F(16, 180) = 1.53$, *ns*.

Research has shown that time judgments obtained under concurrent-task conditions typically conform to power functions (e.g., see Hemmes et al., 2004) that are better analyzed in terms of log-log functions, as opposed to the linear analysis described above. Therefore, the data were transformed to log estimates prior to subsequent analyses. The log-transformed data were analyzed by means of a 5 x 5 x 3 (Feedback x Duration x Session) mixed-factorial ANOVA. The analyses yielded a number of significant effects (see Appendix B, Table B9 for all comparisons from this analysis.) including significant effects of duration, $F(4, 180) = 1242.03$, and feedback, $F(4, 45) = 2.83$. As shown in Figure 1, mean time estimates varied dependent upon type of feedback. Pairwise tests (Fisher's Protected *t*) revealed that mean estimates for the binary group, were significantly lower than those of the magnitude group, $t(45) = 2.78$, the directional group, $t\ (45) = 2.38$, and the combined group, $t\ (45) = 2.42$. No differences in time estimates were found in comparing performance among any other conditions.

Type of feedback also resulted in differences in mean time judgments relative to veridical time. Single sample *t* tests revealed no difference in comparing mean time judgments to veridical time for the magnitude group, $t(45) = .29$, *ns*; the directional group, t (45) = .87, *ns;* or the combined group, t (45) = .80, *ns*. In contrast, mean temporal judgments were significantly shorter than veridical time for the binary group, t (45) = 3.80, and the no-feedback group, $t(45) = 2.91$.

In addition to the significant effects of both duration and feedback, the analyses revealed a significant difference in mean log time judgments across experimental sessions, $F(2, 90) = 5.73$. As shown in Figure 2, it appears that judgments were shorter than veridical time across the three experimental sessions, and that time judgments were lower in the first session than in the remaining two sessions. Single-sample t tests confirmed that judgments were significantly shorter than veridical during all experimental sessions. Differences in mean time judgments across sessions were assessed using contrasts. The results of these analyses revealed no differences in performances during sessions in which feedback was presented (Sessions 1 & 2) in comparison to the session during which feedback was absent (Session 3), $F(1, 90) = 1.15$, *ns*. Moreover, no difference in mean time judgments was observed in comparing performance during the two sessions (Sessions 1 & 2) in which feedback was presented, $F(1, 90) = 1.14$, *ns*.

Figure 3 depicts mean time judgments as a function of sample duration, separately for each feedback group. The solid, diagonal line indicates veridical time. As shown in the graph, for all feedback groups, time judgments were direct functions of sample duration. Regression analysis of group mean judgments revealed that all functions conformed to a linear relationship ($r^2 > .99$ for all feedback functions). Single-sample t tests comparing the slopes of each of the feedback functions with veridical time confirmed that all functions significantly differed from veridical time ($p < .05$).

As shown in Figure 3, there is some separation of the functions dependent upon type of feedback. Moreover, the slope of the function for the no-feedback group appears flatter than those obtained for the groups receiving some form of feedback. Nevertheless, contrary to predictions, and consistent with the analysis of the untransformed estimates, the Feedback x Duration interaction failed to reach significance, $F(16, 180) = 2.08$, *ns*. Analysis of the slope values obtained for each participant in each feedback group failed to reach significance, $F(4, 45) = 2.17$, *ns*, suggesting that all functions are parallel. Differences in intercept values also failed to reach significance, $F(1, 90) = 1.64$, *ns*.

Variability of Temporal Judgments

Whereas the preceding analyses show that type of feedback affects the accuracy of log time judgments, also of interest was the impact of feedback on the variability in time judgments. The *SD*s of the untransformed estimates were calculated for each participant in each feedback group for every combination of experimental session and sample duration. A constant value of 1 was added to these data because 27 measures (3.6% of the data set) had a value of zero, and thus, log transformations were not possible for these measures. The *SD*s were then log transformed, and subsequently analyzed by means of a 5 x 5 x 3 (Feedback x Duration x Session) mixed-factorial ANOVA. The results of these analyses are presented in Appendix B, Table B10.

Although type of feedback did not have a significant main effect on variability of log estimates, $F(4, 45) = 1.08$, *ns*, the analysis revealed a significant effect of duration, $F(4, 180) = 103.18$, and a significant Feedback x Duration interaction, $F(16, 180) = 2.04$. The relation between feedback, duration and variability is shown in Figure 4 which indicates that *SD* of time judgments increased with duration for all feedback groups;

however, it appears that the amount of increase in variability depended upon both sample duration and type of feedback. The data show that between 5 and 8 s, there is a greater increase in variability of time judgments for the magnitude, directional and combined feedback groups relative to both the binary and no-feedback groups. For sample durations of at least 8 s, increases in variability appear similar across all feedback groups with one exception; there is a decrease in variability from 17 to 25 s for the directional feedback group.

To allow for a more thorough interpretation of the interaction effect, one-way ANOVAs were conducted on the variability measures individually for each feedback group. The effect of duration on variability measures was significant for all feedback groups (minimum: $F(4, 36) = 2.74$). Pairwise testing was conducted using Tukey's HSD. A complete summary of the results of these comparisons is presented in Appendix B, Table B11. Consistent with the interpretation of the graphical representation of the data, the results of these tests revealed that timing performance was significantly more variable at 8 s than at 5 s for each of the magnitude, directional and combined feedback groups. No such difference in variability was found at these sample durations for either the binary or no-feedback groups. The pairwise tests also revealed that comparisons of performance at 8 with 11 s revealed no difference in variability for any feedback group. As shown in the table, the remainder of the comparisons suggests a complicated pattern of differences in variability measures across duration relative to feedback condition. It does appear, however, that in general, for the no-feedback, binary, magnitude and combined groups, the majority of comparisons of variability measures significantly differed across sample duration. In contrast, for the directional group, differences in variability measures were observed in only half of the comparisons. Moreover, no differences were observed for this group in comparing variability measures at the three longest values of sample duration with one another.

To better understand differences in variability in responding across the five feedback groups, mean log time judgments for each participant at each value of sample duration were plotted separately for each feedback group. These data are presented in Figure 5. In all graphs, linear regression is presented as a solid line; group mean judgments as a dotted line; and veridical time as a dashed line. As shown in the figure, participants in the combined group performed similarly and near veridically at all values of sample duration. In contrast, performance of participants in both the no-feedback and binary groups differed greatly at all values of sample duration. Time judgments provided by participants in the magnitude group also differed across all values of sample duration, but not to the same degree as those in both the no-feedback and binary groups. With respect to the directional group, performance of the individual participants was similar at the shortest values of sample duration (i.e., 5 & 8 s), with greater differences observed at 11 through 25 s.

Although as reported previously, r^2 values based on group means were quite high ($r^2 > .99$ for all feedback functions—see Figure 3), similar regression analysis of individual participant means produced different results (see Table 3). Whereas the r^2 values for the magnitude, directional and combined groups remained high ($r^2 > .90$ for all), there was a marked decrease in the value of r^2 for the no-feedback group ($r^2 = .57$) and the binary group ($r^2 = .73$), suggesting greater heterogeneity in performance of these two groups than the three groups receiving more precise feedback.

Further analysis of the variability measure also revealed a significant effect of session, $F(2, 90) = 9.18$, $p < .05$. As shown in the Figure 6, variability decreased across experimental sessions. Difference in variability of temporal judgments across experimental session was assessed by means of contrasts. No significant difference in variability was detected in comparing performance during sessions in which feedback was presented (Sessions 1 & 2) with performance in the absence of feedback (Session 3), $F(1, 90) = 3.17$, *ns*. In addition, measures of variability did not differ between the two sessions during which feedback was presented, $F(1, 90) = .50$, *ns*.

Discussion

The primary goal of the present study was to test the effects of a number of forms of feedback in the context of a temporal estimation task. An attempt was made to determine how differences in specific properties of these feedback stimuli affect time estimation. It was proposed the type of feedback manipulation might be characterized in terms of a continuum of *precision*. Two indices for determining the precision of feedback were presented, a *feedback-statements model* and a *response-directional-change model*. Although both models predict superior performance given any form of feedback relative to conditions of no-feedback, these models differentially predict performance dependent upon type of feedback.

Precision under the feedback-statements model (see Table 1) is based on the number of feedback statements available given an incorrect estimate. Precision of feedback increases directly with the number of feedback statements contained within a given feedback stimulus. Alternatively, the response-directional-change model (see Table 2) proposed that precision of feedback increases inversely with the number of available response-directional change options following an incorrect estimate of duration; fewer available response-directional change options are indicative of greater precision of the feedback stimulus.

Analyses performed on the time judgments gathered in the present study provide some support of a feedback-statements model of precision of feedback. For example, time judgments for the binary group differed in comparison to those obtained for the three feedback groups in which feedback was characterized as more precise on the basis of number of feedback statements. Upon examination of time judgments pooled across both session and values of sample duration, the binary group produced significantly shorter estimates than those produced by groups exposed to feedback characterized as more precise. Additional support for the feedback-statements model is evident in that there were no differences in performance between the magnitude group and the directional group. Because each of these feedback stimuli contains the same number of feedback statements, the model predicts similar performance under these two conditions. In general, the outcomes of the analyses confirm this prediction.

One shortcoming of the feedback-statements model arises with respect to data of the combined group. Under this feedback condition, all three feedback statements were provided, so the stimuli presented to participants in this group were the most precise as indexed by the model. In general, the analyses failed to reveal superior performance of the combined group in comparison to both the magnitude and directional groups. It is important to note that the lack of difference in performance across feedback groups is likely attributable to a ceiling effect in that time estimates did not differ from veridical time for these three groups (see Figure 1).

Although the data do lend some support to a feedback-statements model of precision, analyses yielded minimal evidence in support of the response-directional-change model. Even though the analyses revealed that timing performance was superior for groups receiving feedback specified as most precise by this model—directional and combined—relative to the less precise binary group, there is more evidence in opposition than in support of response-directional-change model. For example, the results of all analyses showed no significant differences in temporal judgments when comparing performance of the magnitude group to that of either of the groups receiving the most precise feedback; namely, the directional and combined groups. Moreover, the response-directional-change model predicts similar performance for the binary and magnitude groups. This was not the case; temporal judgments provided by the magnitude group were significantly more accurate than those provided by the binary group.

Regardless of the fact that the data obtained in the present study did not conform entirely to either of the models proposed herein, possibly owing to a ceiling effect, it is important to note that type of feedback did differentially affect timing performance. The log-transformed data showed differences in timing performance relative to the type of feedback. When pooled across all values of sample duration and experimental session, mean log estimates were significantly lower for the binary group, for which feedback referred only to response accuracy versus inaccuracy, in comparison to the groups receiving more feedback, referencing the magnitude, and/or the direction of error. This finding indicates that when feedback was correlated with some property of the discrepancy between physical time and subjective time (beyond that of accuracy), time judgments were not statistically different from veridical time. Because there were no differences observed in comparing performance of the combined group to that of both the magnitude and directional groups, it appears that for feedback to differentially affect timing performance, these stimuli need only correlate with accuracy and one additional property of the difference between a given judgment and sample duration.

Although it was expected that the effect of type of feedback might vary in relation to sample duration owing to the greater degree of underestimation typically observed at the longer as opposed to shorter values of sample duration, the obtained *p* value (.0544) was slightly below the criterion for significance (.05) adopted for this study.

Accordingly, the effect of type of feedback on mean temporal judgments did not depend on the value of sample duration. Further analyses revealed that the slopes of the functions for the five feedback groups (refer to Figure 3) were linear, and that the slopes of these functions did not differ statistically from one another, suggesting that the functions for all feedback groups are approximately parallel. In accord with Stevens' (1960) observations regarding responding to changes in physical stimulus dimensions, the similarity in slopes of log-log functions suggests equal sensitivity to changes in stimulus duration across feedback groups. As described previously, the analysis revealed that time judgments were shorter under the binary condition than the other feedback conditions. Taken together, these findings suggest that type of feedback affects the accuracy of temporal judgments, as opposed to sensitivity to duration. In addition, because of the lack of differences among the slopes of the functions, and the findings that judgments under the binary condition were consistently lower than all other conditions across the range of durations, there is evidence to suggest that type of feedback affected accuracy of estimates by a constant as measured across a range of durations.

One interesting finding of the preceding analysis pertains to the data for the no-feedback group. As stated previously, analysis of the individual slopes of the functions for each feedback group (see Figure 3) revealed no significant difference in slope measures even though upon visual inspection, these data suggest that the slope of the no-feedback function is flatter than the slopes the functions of the groups who received some form of feedback. Moreover, at all values of sample duration with the exception of 25 s, judgments of the no-feedback group appear closer to veridical time than those provided by the binary group. When time judgments for each feedback group were collapsed across experimental session and sample duration, analysis yielded no differences in temporal judgments for the no-feedback group compared with the groups receiving some form of feedback. Furthermore, for both the no-feedback and binary groups, the analysis revealed that time judgments were significantly shorter than veridical time. Based on the preceding, it appears that merely providing feedback related to accuracy (i.e., "correct" or "incorrect") has a negative impact on timing performance; participants not receiving feedback outperformed those receiving binary feedback.

As described in the preceding paragraphs and contrary to predictions, the effect of type of feedback on time judgments did not differ dependent upon sample duration. Nonetheless, similar analysis performed on measures of variability did reveal a significant interaction of feedback with sample duration. The data suggested that the interaction effect was most prominent in comparing variability in temporal judgments at the two shortest values of sample duration. Testing revealed that variability measures for both the no-feedback group, and the least precise feedback group—the binary group—did not differ in comparing performance at 5 and 8 s. In contrast, at theses values of sample duration, significant differences in variability were observed for the more precise feedback groups. Moreover, at 8 s and beyond, a steady and similar increase in variability of time judgments was observed regardless of feedback group. It is important to note that these findings contradict the predictions made in the introduction. There was an expectation that consistent with previous research (e.g., Allen & Clark, 1979; Hicks & Miller, 1976; Montare, 1988; & Robinson, 1963), variability would decrease with the presentation of some form of feedback.

In addition to the effects of feedback described above, the analysis of the log-transformed estimates also yielded a significant effect of session. Although graphical representation of these data show that mean temporal judgments increased across sessions (see Figure 2), contrasts revealed no differences in performance during sessions with (Sessions 1 & 2) versus the session without (Session 3) feedback. Moreover, no differences in temporal judgments were found in comparing performance during the first two session during which feedback was presented.

The finding that mean time judgments did not significantly change upon removal of feedback during the final test session is in line with the findings of Hicks and Miller 1976), yet contrary to the results reported by Allen and Clark (1979) as well as predictions made in the present study. Hicks an Miller found that when participants were exposed to an estimation task across a range of durations, when feedback was presented during the initial phase of testing, removal of feedback in a second phase did not cause differences in judgment errors regardless of whether participant were exposed to the same or different ranges of durations across the two experimental phases.

In contrast to the findings of Allen and Clark (1979), in the present study, no differences in time judgments were observed when participants who were exposed to feedback were subsequently tested in the absence of feedback. Although Montare (1988) reported no significant difference in the effect of feedback compared across two methods of temporal judgment (i.e., production vs. estimation), it is possible that the difference regarding the short-term effects of feedback reported by Allen and Clark in comparison to those obtained in the present research may be due to methodological differences and/or differences related to type of dependent measure. Allen and Clark presented feedback across a range of durations during an estimation task, and then using a temporal discrimination task, tested for short-term effects of feedback at a single duration selected from the range used during the estimation task. In the present study, both values of duration, and type of timing task, were held constant while feedback was manipulated between present and absent. An additional methodological difference between the present study and the aforementioned research is that in the present study, a concurrent, number-reading task was presented to interfere with chronometric counting. Although, participants in the Hicks and Miller study were instructed to refrain from counting, in neither their study nor that performed by Allen and Clark was an additional task added in an attempt to interfere with counting the duration value.

Changes in performance across sessions were observed not only with respect to the accuracy of temporal judgments, but also with respect to measures of variability. The analysis revealed a significant effect of session on measures of variability. Although visual examination of these data suggested a decrease in variability across sessions, contrasts revealed no differences in variability of time judgments across sessions. Nonetheless, decrease in variability across sessions is in line with the findings reported by both Robinson (1963) and Allen and Clark (1979) who found decreases in variability with repeated exposure to feedback.

As described in the preceding paragraphs, the results of the study suggest that timing performance differs dependent upon type of feedback. Although these findings are interesting, it is important to again mention that the primary goal of the present study was to attempt to provide a metric for scaling the level of precision of a variety of forms of feedback stimuli. Although there is no conclusive evidence in support of either of the two indices characterizing the precision feedback stimuli, the results of the analyses suggest that perhaps, for feedback stimuli that contain no more than two feedback statements, the feedback-statements model provides a valid index of the precision of feedback stimuli. Moreover, a stimulus containing two feedback statements (i.e., magnitude & directional feedback) was sufficient to induce near-veridical timing. This would suggest that the third feedback statement contained in the combined feedback stimuli had no additional function in supporting responding. An alternative explanation centers on ceiling effects. It is possible that the maximum effects of these three types of feedback were observed across the range of sample durations used in the present study. To test the difference between these two interpretations, one might assess the effect of these types of feedback stimuli across a greater number of sample durations and/or larger values of sample duration. Because timing under concurrent-task conditions typically results in increases in underestimation with increases in duration (e.g., see Hemmes et al. 2004), it is possible that at longer values of sample duration, a ceiling effect might be avoided, and differential effects of type of feedback, observed.

Although the results of the present study provide no evidence clearly in favor of either of the two models of precision of feedback presented herein, it is plausible that this research will provide a platform for further investigation addressing the implications of feedback in human timing studies. Ideally, the interpretations of the findings of the present study will not only provided a basis for further timing research, but in studies of other perceptual processes.

Footnotes

[1] In most feedback applications, partial feedback refers to an intermittent schedule of feedback (e.g., see Kladopoulos et al., 1998) as opposed to Robinson's (1963) use referring to differences in stimuli among feedback conditions.

[2] In post-hoc interviews with the participants in the present study as well as in pilot research, a few reported that they had inadvertently typed an extra number (e.g., "89" as opposed to "8") when entering their estimate of a duration. Therefore, estimates that were 4 standard deviations or more beyond the mean estimate for a given duration (pooled across feedback condition) were not included in the analyses in the present study.

[3] The feedback-statements model (see Table 1) proposes that magnitude and directional feedback are equally precise. The order of these two conditions as scaled in Figure 1 is arbitrary.

Table 1

Feedback Statements Model of Precision of Feedback

Feedback Condition	Incorrect	Off by *t*	Too short/Too long	Total Statements
No-feedback				0
Binary	1			1
Magnitude	1	1		2
Directional	1		1	2
Combined Directional and Magnitude	1	1	1	3

Note. This model proposes that the precision of a feedback stimulus increases as the total number of feedback statements contained in a feedback stimulus increases. This analysis assumes a given temporal stimulus and an incorrect estimate, and applies to all stimulus/incorrect estimate pairs. In the magnitude, directional and combined conditions, the incorrect feedback statement is represented by either the words "off by" or "too."

Table 2

Response-directional-change Model of Precision of Feedback

Feedback Condition	No change	Increase	Decrease	Available Responses
No-feedback	1	1	1	3
Binary		1	1	2
Magnitude		1	1	2
Directional		1	1	1
Combined Directional and Magnitude		1	1	1

Note. The model proposes that a reduction in response directional changes leads to increases in the precision of feedback stimuli. The analyses refer to directional responses available given an incorrect estimate, assuming that the participant attempts to comply with the feedback. In the directional and combined conditions, if there is compliance with the feedback, then a participant may *either* increase or decrease the estimate, but both response options are not available simultaneously as is the case in the binary and magnitude conditions.

Table 3

Summary of Regression Analysis of Individual Participant Mean Log Estimates across the Five Sample Durations for each Feedback Group

Feedback Group	slope	y intercept	r^2
No Feedback	.69	.24	.5689
Binary	.76	.13	.7326
Magnitude	.80	.20	.9307
Directional	.87	.09	.9488
Combined	.83	.16	.9867

Note. The slope and intercept values presented in the table are the values obtained for the regression line for each feedback group.

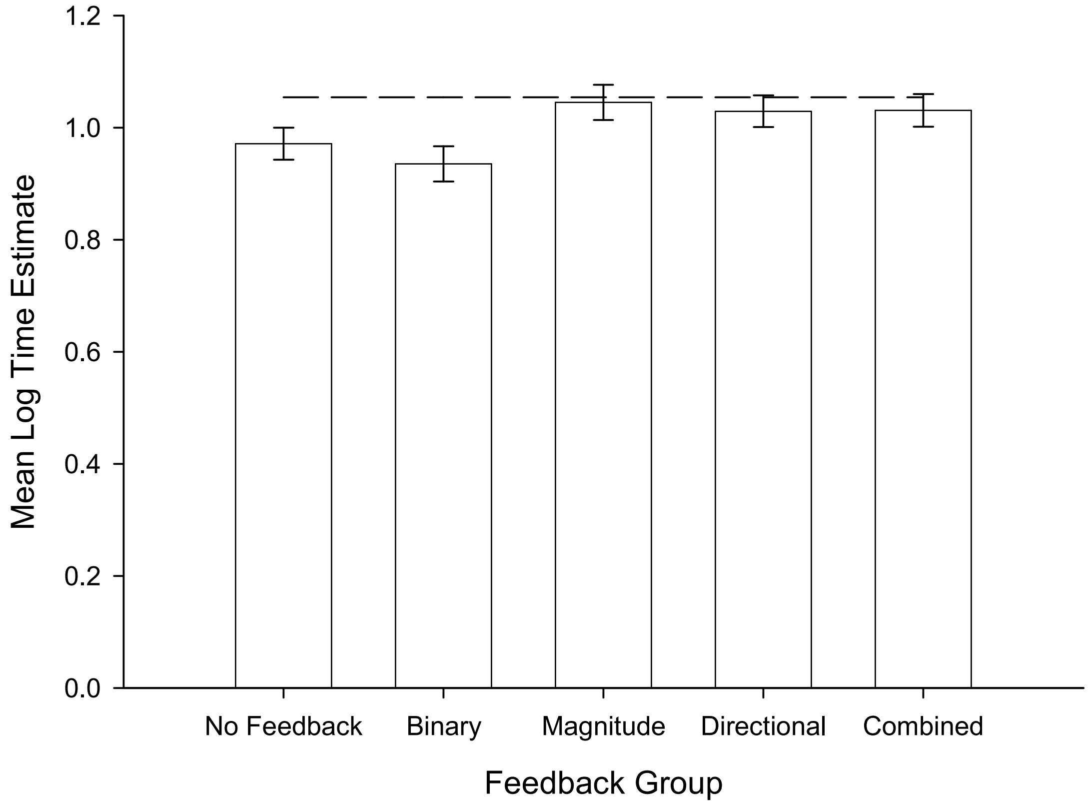

Figure 1. Mean log time estimates and standard error of mean log time estimates for each feedback group. The five feedback conditions are presented on the abscissa and are scaled from least to most precise according to the feedback-statements model (see Table 1)[3]. The data are collapsed across sample duration and session. Veridical time is represented by a horizontal, dashed line.

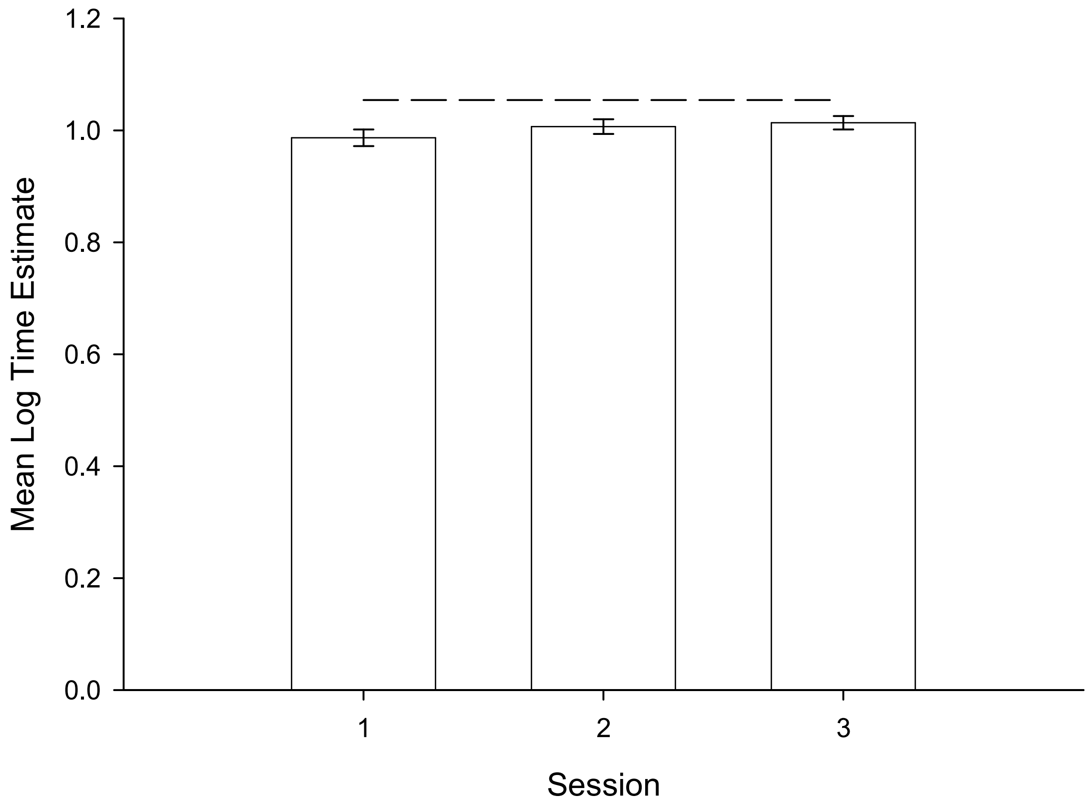

Figure 2. Mean log time estimates and standard error of mean log time estimates for each of the three experimental sessions. The data are collapsed across sample duration and feedback group. Veridical time is represented by a horizontal, dashed line.

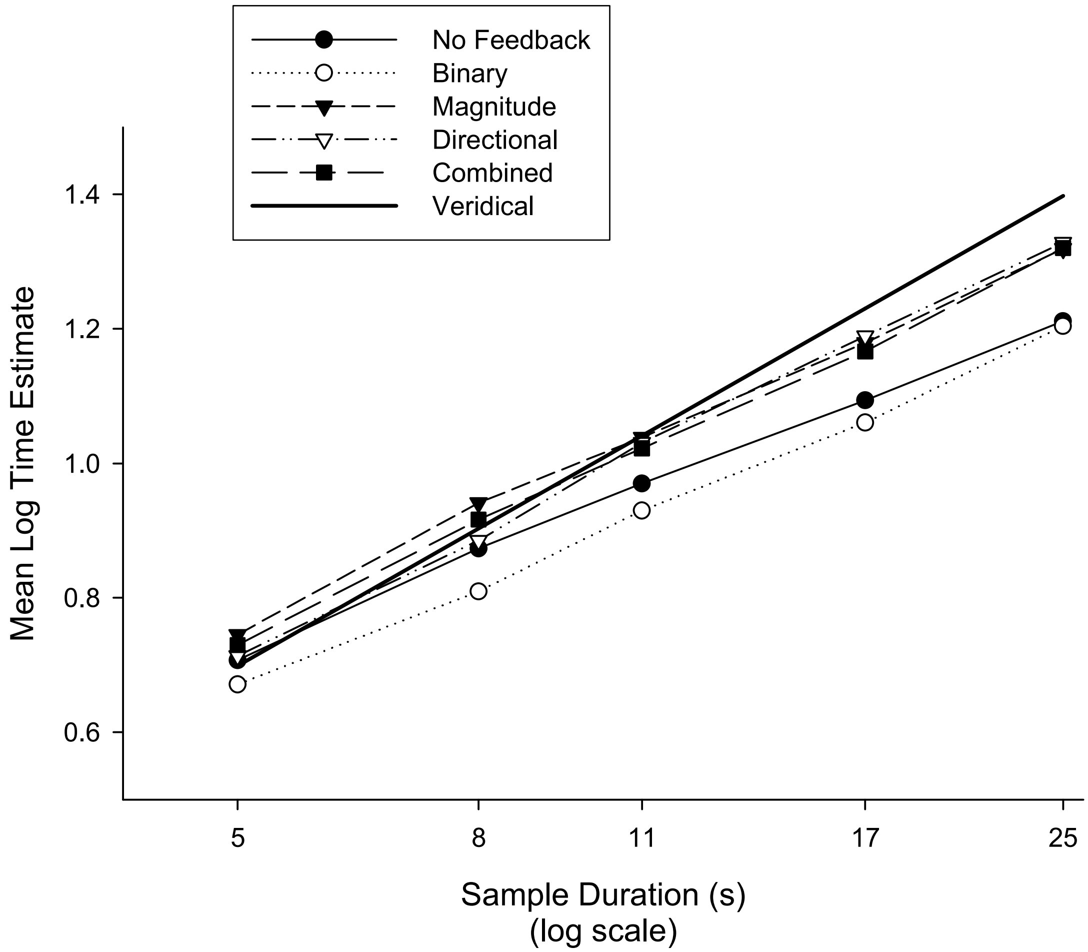

Figure 3. Mean log time estimates as a function of log sample duration for each of the five feedback groups. The solid, diagonal line depicts log veridical time.

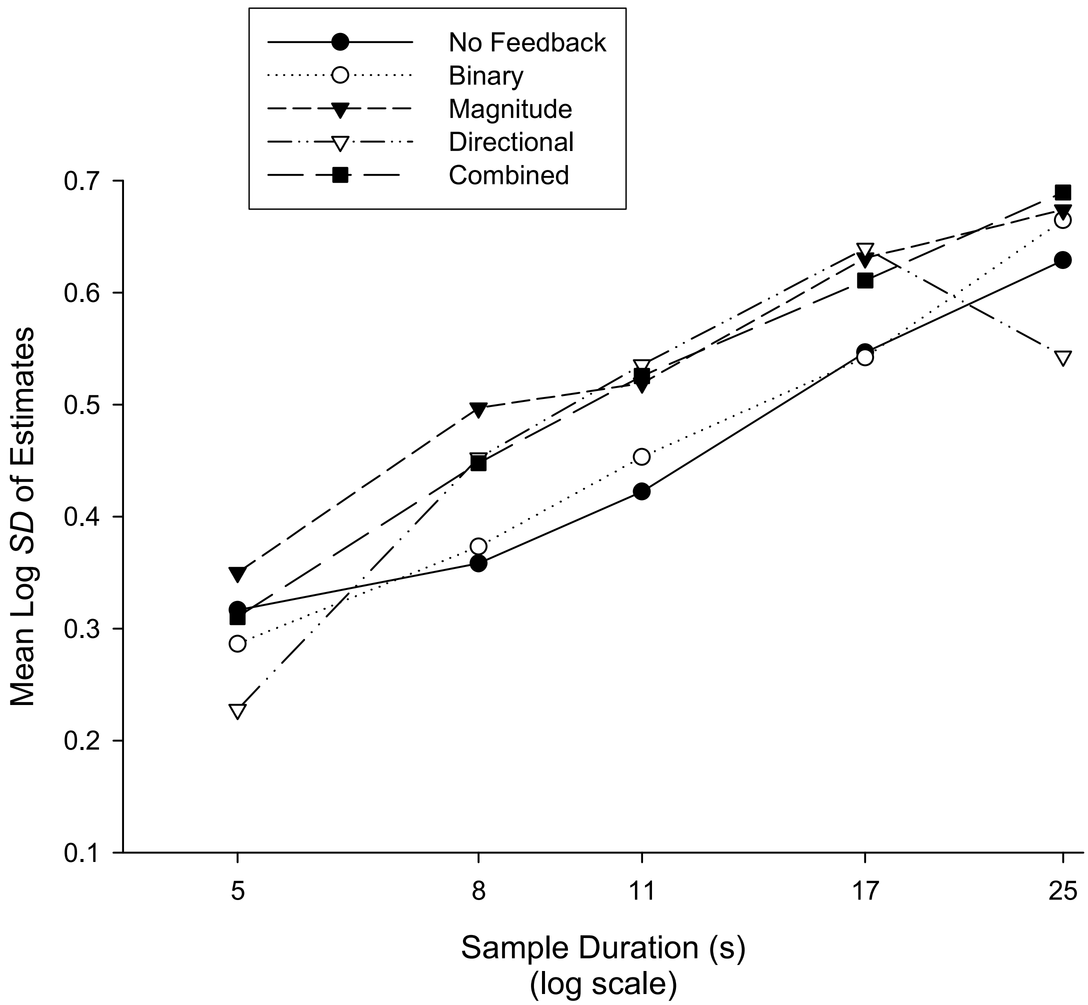

Figure 4. Mean log *SD* of untransformed time estimates as a function of log sample duration for each of the five feedback groups.

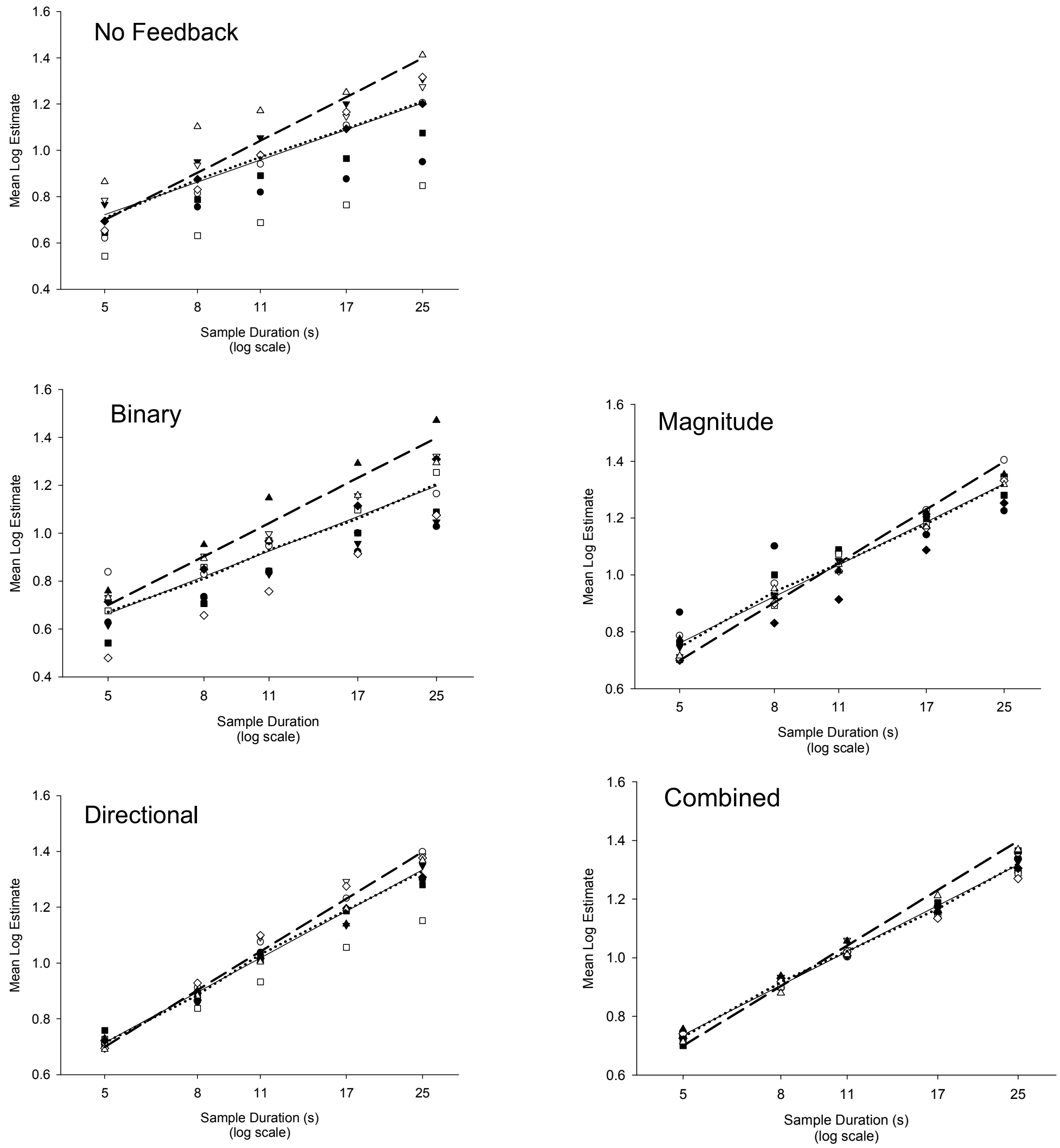

Figure 5. The panels show mean log time estimates for each participant (N = 10 per feedback group) at each duration under the individual feedback conditions. In each panel, linear regression is represented by the solid line. Veridical time is represented by the dashed line. Group mean estimates are represented by the dotted line.

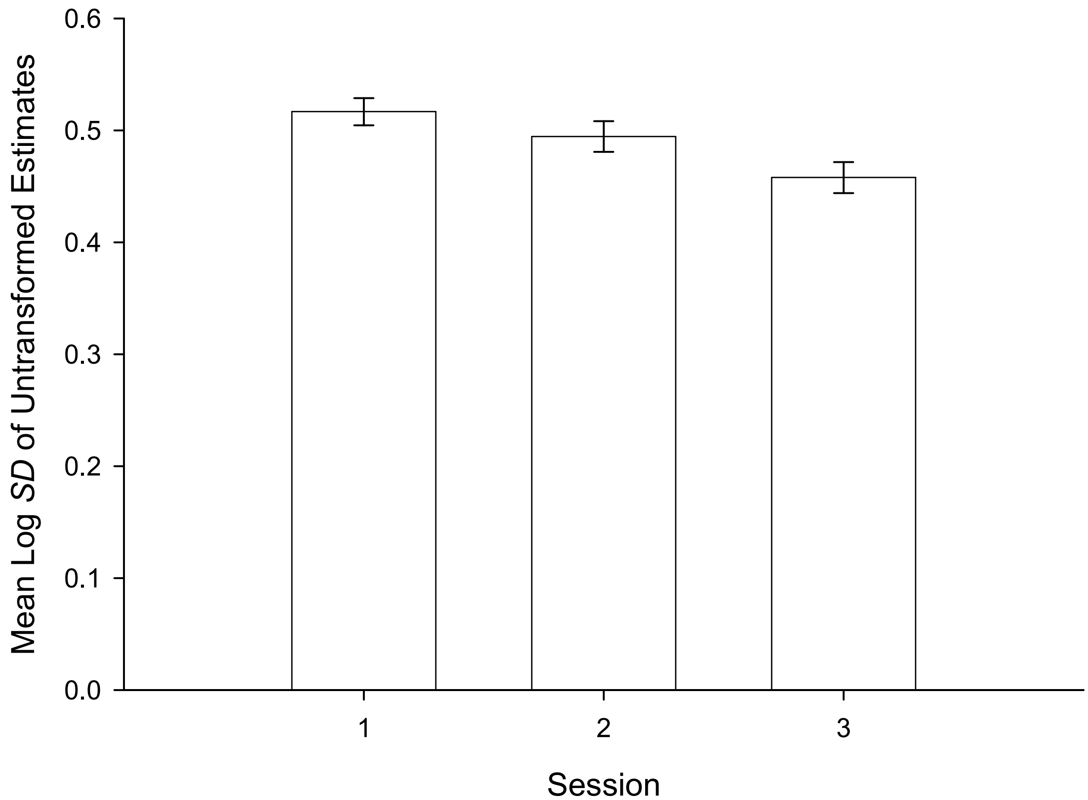

Figure 6. Mean log *SD* of untransformed time estimates and standard error of mean log *SD* for each of the three experimental sessions. The data are collapsed across sample duration and feedback group.

Appendix A

No-feedback Condition.

INSTRUCTIONS FOR TEST SESSIONS

On each trial you will see a GREEN SQUARE on the computer for a specific amount of time. We will call that the SAMPLE time. When the GREEN SQUARE disappears, you will be asked to ESTIMATE THE AMOUNT OF TIME OF THE GREEN SQUARE was on the screen to the nearest second.

Please type in your estimate using the NUMBER KEYS and then press ENTER.

PLEASE BE CAREFUL! Once you press enter, you CANNOT go back and change your estimate of the SAMPLE time.

On ALL trials, numbers will flash at the center of the GREEN SQUARE. You are to read each number aloud to its full place name.
For example, 4692 should be read as four thousand, six hundred and ninety-two. Your voice will be recorded.

Remember, your task is to ESTIMATE the SAMPLE time. That is, the amount of time the GREEN SQUARE remained on the screen.

NO FEEDBACK will be provided. After each ESTIMATE, you will see a message to WAIT. Press the space bar for the next trial AFTER this message disappears.

You will see a message to call the experimenter when the session is over.

----------STOP AT THIS POINT AND CALL THE EXPERIMENTER----------

----------REPEAT INSTRUCTIONS TO EXPERIMENTER----------

DO YOU HAVE ANY QUESTIONS???

If you don't, then....

Binary Feedback Condition.

INSTRUCTIONS FOR TEST SESSIONS

On each trial you will see a GREEN SQUARE on the computer for a specific amount of time. We will call that the SAMPLE time. When the GREEN SQUARE disappears, you will be asked to ESTIMATE THE AMOUNT OF TIME OF THE GREEN SQUARE was on the screen to the nearest second.

Please type in your estimate using the NUMBER KEYS and then press ENTER.

PLEASE BE CAREFUL! Once you press enter, you CANNOT go back and change your estimate of the SAMPLE time.

On ALL trials, numbers will flash at the center of the GREEN SQUARE. You are to read each number aloud to its full place name.
For example, 4692 should be read as four thousand, six hundred and ninety-two. Your voice will be recorded.

Remember, your task is to ESTIMATE the SAMPLE time. That is, the amount of time the GREEN SQUARE remained on the screen.

FEEDBACK will be provided following each trial. After each ESTIMATE is entered, you will see a message telling you whether your ESTIMATE of the SAMPLE was CORRECT or INCORRECT. You may use this information to help you with future judgments.

The FEEDBACK will disappear after a certain amount of time.

Press the space bar for the next trial AFTER the FEEDBACK disappears.

You will see a message to call the experimenter when the session is over.

----------STOP AT THIS POINT AND CALL THE EXPERIMENTER----------

----------REPEAT INSTRUCTIONS TO EXPERIMENTER----------

DO YOU HAVE ANY QUESTIONS???

If you don't, then....

Magnitude Feedback Condition.

INSTRUCTIONS FOR TEST SESSIONS

On each trial you will see a GREEN SQUARE on the computer for a specific amount of time. We will call that the SAMPLE time. When the GREEN SQUARE disappears, you will be asked to ESTIMATE THE AMOUNT OF TIME OF THE GREEN SQUARE was on the screen to the nearest second.

Please type in your estimate using the NUMBER KEYS and then press ENTER.

PLEASE BE CAREFUL! Once you press enter, you CANNOT go back and change your estimate of the SAMPLE time.

On ALL trials, numbers will flash at the center of the GREEN SQUARE. You are to read each number aloud to its full place name.
For example, 4692 should be read as four thousand, six hundred and ninety-two. Your voice will be recorded.

Remember, your task is to ESTIMATE the SAMPLE time. That is, the amount of time the GREEN SQUARE remained on the screen.

FEEDBACK will be provided following each trial. After each ESTIMATE is entered, if your estimate is incorrect, you will see a message telling you
by how many seconds your ESTIMATE is incorrect. You may use
this information to help you with future judgments.

The FEEDBACK will disappear after a certain amount of time.

Press the space bar for the next trial AFTER the FEEDBACK disappears.

You will see a message to call the experimenter when the session is over.

----------STOP AT THIS POINT AND CALL THE EXPERIMENTER----------

----------REPEAT INSTRUCTIONS TO EXPERIMENTER----------

DO YOU HAVE ANY QUESTIONS???

If you don't, then....

Directional Feedback Condition.

INSTRUCTIONS FOR TEST SESSIONS

On each trial you will see a GREEN SQUARE on the computer for a specific amount of time. We will call that the SAMPLE time. When the GREEN SQUARE disappears, you will be asked to ESTIMATE THE AMOUNT OF TIME OF THE GREEN SQUARE was on the screen to the nearest second.

Please type in your estimate using the NUMBER KEYS and then press ENTER.

PLEASE BE CAREFUL! Once you press enter, you CANNOT go back and change your estimate of the SAMPLE time.

On ALL trials, numbers will flash at the center of the GREEN SQUARE. You are to read each number aloud to its full place name.
For example, 4692 should be read as four thousand, six hundred and ninety-two. Your voice will be recorded.

Remember, your task is to ESTIMATE the SAMPLE time. That is, the amount of time the GREEN SQUARE remained on the screen.

FEEDBACK will be provided following each trial. After each ESTIMATE is entered, you will see a message telling you whether your ESTIMATE of the SAMPLE was CORRECT, TOO SHORT or TOO LONG. You may use this information to help you with future judgments.

The FEEDBACK will disappear after a certain amount of time.

Press the space bar for the next trial AFTER the FEEDBACK disappears.

You will see a message to call the experimenter when the session is over.

----------STOP AT THIS POINT AND CALL THE EXPERIMENTER----------

----------REPEAT INSTRUCTIONS TO EXPERIMENTER----------

DO YOU HAVE ANY QUESTIONS???

If you don't, then....

Combined Feedback Condition.

INSTRUCTIONS FOR TEST SESSIONS

On each trial you will see a GREEN SQUARE on the computer for a specific amount of time. We will call that the SAMPLE time. When the GREEN SQUARE disappears, you will be asked to ESTIMATE THE AMOUNT OF TIME OF THE GREEN SQUARE was on the screen to the nearest second.

Please type in your estimate using the NUMBER KEYS and then press ENTER.

PLEASE BE CAREFUL! Once you press enter, you CANNOT go back and change your estimate of the SAMPLE time.

On ALL trials, numbers will flash at the center of the GREEN SQUARE. You are to read each number aloud to its full place name.
For example, 4692 should be read as four thousand, six hundred and ninety-two. Your voice will be recorded.

Remember, your task is to ESTIMATE the SAMPLE time. That is, the amount of time the GREEN SQUARE remained on the screen.

FEEDBACK will be provided following each trial. After each ESTIMATE is entered, you will see a message telling you by how many seconds your estimate is incorrect [Zero (0) seconds = correct estimate].

You will also see a message telling you whether your ESTIMATE of the SAMPLE was CORRECT, TOO SHORT or TOO LONG. You may use this information to help you with future judgments.

The FEEDBACK will disappear after a certain amount of time.

Press the space bar for the next trial AFTER the FEEDBACK disappears.

You will see a message to call the experimenter when the session is over.

----------STOP AT THIS POINT AND CALL THE EXPERIMENTER----------

----------REPEAT INSTRUCTIONS TO EXPERIMENTER----------

DO YOU HAVE ANY QUESTIONS???

If you don't, then....

Appendix B

Table B1

Mean Time Estimates and Standard Deviation of Mean Estimates for each Feedback Group

Feedback Group	Mean (s)	*SD*
No Feedback	11.18	7.30
Binary	10.19	6.57
Magnitude	12.77	6.23
Directional	12.50	6.87
Combined	12.28	6.37

Note. The values presented in the table represent the data collapsed across session and sample duration. Veridical time collapsed across the five values of sample duration is 13.20 s.

Table B2

Mean Time Estimates and Standard Deviation of Mean Estimates for each Experimental Session

Session	Mean (s)	*SD*
Session 1	11.41	6.69
Session 2	11.90	7.00
Session 3	12.05	6.75

Note. The values presented in the table represent the data collapsed across feedback group and sample duration. Veridical time collapsed across the five values of sample duration is 13.20 s.

Table B3

Mean Time Estimates and Standard Deviation of Mean Estimates at each Value of Sample Duration

Sample Duration (s)	Mean (s)	*SD*
5	5.38	1.59
8	8.12	2.83
11	10.54	4.12
17	14.64	4.92
25	20.24	6.74

Note. The values presented in the table represent the data collapsed across feedback group and session.

Table B4

Mean Time Estimates and Standard Deviation of Mean Estimates for each Feedback Group across each of the three Experimental Sessions

Feedback Group	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
No Feedback	11.37	7.72	11.39	7.44	10.80	6.89
Binary	9.85	6.36	10.22	7.15	10.48	6.14
Magnitude	11.78	6.31	12.78	6.58	13.77	6.86
Directional	12.14	6.68	12.69	6.90	12.69	7.01
Combined	11.86	6.04	12.41	6.58	12.55	6.46

Note. The values presented in the table represent the data collapsed across sample duration. Veridical time collapsed across the five values of sample duration is 13.20 s.

Table B5

Mean Time Estimates and Standard Deviation of Mean Estimates for each Feedback Group at each Value of Sample Duration

Sample Duration (s)	No-feedback		Binary Feedback		Magnitude Feedback	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	5.37	1.79	4.96	1.64	5.81	1.92
8	8.03	3.11	6.84	2.28	9.28	3.61
11	10.07	3.99	9.26	6.04	11.39	3.42
17	13.81	6.29	12.37	4.58	15.74	4.35
25	18.64	9.53	17.57	7.25	21.53	4.95

Sample Duration (s)	Directional Feedback		Combined Feedback	
	Mean (s)	*SD*	Mean (s)	*SD*
5	5.26	1.04	5.50	1.27
8	7.97	2.24	8.47	2.05
11	11.19	3.40	10.81	2.55
17	16.16	4.46	15.07	3.54
25	21.90	4.69	21.51	4.71

Note. The values presented in the table represent the data collapsed across experimental session.

Table B6

Mean Time Estimates and Standard Deviation of Mean Estimates for each Experimental Session at each Value of Sample Duration

Sample Duration (s)	Experimental Session					
	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	5.21	1.63	5.42	1.48	5.51	1.63
8	7.99	2.76	8.03	2.80	8.34	2.92
11	10.10	3.35	10.77	5.18	10.75	3.56
17	13.90	4.82	14.87	4.99	15.13	4.88
25	19.83	7.11	20.39	6.66	20.50	6.41

Note. The values presented in the table represent the data collapsed across feedback group.

Table B7

Mean Time Estimates and Standard Deviation of Mean Estimates for each Feedback Group for each Experimental Session at each Value of Sample Duration

Sample Duration (s)	No Feedback					
	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	5.38	1.93	5.45	1.82	5.28	1.63
8	8.30	3.39	8.06	3.05	7.72	2.87
11	9.97	3.73	10.17	4.12	10.06	4.13
17	13.63	5.96	14.35	6.89	13.47	5.99
25	19.56	10.73	18.98	9.34	17.41	8.31
Sample Duration (s)	Binary Feedback					
	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	4.71	1.57	5.13	1.69	5.04	1.63
8	6.67	2.24	6.69	2.08	7.18	2.30
11	8.99	3.95	9.52	9.31	9.27	2.79
17	11.78	5.15	12.59	4.36	12.73	4.17
25	17.24	7.61	17.27	7.08	18.17	7.07
Sample Duration (s)	Magnitude Feedback					
	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	5.34	1.83	5.81	1.66	6.29	2.12
8	8.40	2.91	9.20	3.64	10.25	3.97
11	10.38	3.05	11.43	3.32	12.36	3.61
17	14.40	4.22	15.67	3.90	17.16	4.49
25	20.36	4.97	21.59	5.05	22.65	4.61
Sample Duration (s)	Directional Feedback					
	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	5.30	1.35	5.27	.91	5.22	.79
8	7.92	2.31	8.06	2.32	7.94	2.11
11	10.61	2.95	11.79	3.33	11.17	3.80
17	15.24	4.13	16.30	4.82	16.93	4.27
25	21.57	5.01	22.01	4.54	22.12	4.54
Sample Duration (s)	Combined Feedback					
	Session 1		Session 2		Session 3	
	Mean (s)	*SD*	Mean (s)	*SD*	Mean (s)	*SD*
5	5.32	1.30	5.43	1.06	5.74	1.40
8	8.63	2.21	8.15	1.98	8.63	1.92
11	10.57	2.70	10.98	2.50	10.89	2.46
17	14.42	3.62	15.42	3.49	15.76	3.47
25	20.34	4.74	22.06	4.65	22.13	4.57

Table B8

Analysis of Variance of Time Estimates across Feedback Groups

Source	*df*	*F*		*p*
		Between Subjects		
Feedback	4	1.99		.1121
		Within Subjects		
Session	2	5.25	**	.0096
Duration	4	468.83	**	<.0001
Session x Feedback	8	2.01		.0636
Feedback x Duration	16	1.53		.1880
Session x Duration	8	1.66		.1579
Session x Feedback x Duration	32	.99		.4743

Note. Asterisks denote significant effects at the *p* value specified in the table. The *p* values of all within-subjects effects were corrected using the Huynh-Feldt epsilon.

Table B9

Analysis of Variance of Performed on the Log-transformed Estimates across Feedback Groups

Source	*df*	*F*		*p*
		Between Subjects		
Feedback	4	2.83	**	.0355
		Within Subjects		
Session	2	5.73	**	.0092
Duration	4	1242.03	**	<.0001
Session x Feedback	8	1.52		.1818
Feedback x Duration	16	2.08		.0544
Session x Duration	8	1.38		.2249
Session x Feedback x Duration	32	.77		.7507

Note. Asterisks denote significant effects at the *p* value specified in the table. The *p* values of all within-subjects effects were corrected using the Huynh-Feldt epsilon.

Table B10

Analysis of Variance of Standard Deviations of Time Judgments

Source	*df*	*F*		*p*
		Between Subjects		
Feedback	4	1.08		.3767
		Within Subjects		
Session	2	9.18	**	.0003
Duration	4	103.18	**	<.0001
Session x Feedback	8	.34		.9446
Feedback x Duration	16	2.14	**	.0138
Session x Duration	8	1.13		.3409
Session x Feedback x Duration	32	.85		.7014

Note. These analyses were based on the *SD*s of the untransformed estimates. A constant value of 1 was added, and then was followed by log transformations of the *SD* values. Asterisks denote significant effects at the *p* value specified in the table. The *p* values of all within-subjects effects were corrected using the Huynh-Feldt epsilon.

Table B11

A Summary of Pairwise Tests (Tukey's HSD) of the Effect of Duration on Log-transformed SDs for each Feedback Group

Sample Duration		No-feedback	Binary	Magnitude
5 s vs.	8 s	*ns*	*ns*	< .01
	11 s	< .05	< .01	< .01
	17 s	< .01	< .01	< .01
	25 s	< .01	< .01	< .01
8 s vs.	11 s	*ns*	*ns*	*ns*
	17 s	< .01	< .01	< .01
	25 s	< .01	< .01	< .01
11 s vs.	17 s	< .05	*ns*	< .01
	25 s	< .01	< .01	< .01
17 s vs.	25 s	*ns*	< .05	< .01
		Directional	Combined	
5 s vs.	8 s	< .01	< .01	
	11 s	< .01	< .01	
	17 s	< .01	< .01	
	25 s	< .01	< .01	
8 s vs.	11 s	*ns*	*ns*	
	17 s	< .05	< .01	
	25 s	*ns*	< .01	
11 s vs.	17 s	*ns*	*ns*	
	25 s	*ns*	< .01	
17 s vs.	25 s	*ns*	*ns*	

Note. The results presented above reflect the pooled data (log transformed) across the three experimental sessions. The data presented show the level of significance (*p* value) of each *t* test. The *p* values were corrected using the Huynh-Feldt epsilon.

Bibliography

Allen, D.A. & Clark, C.A. (1979). Effect of information feedback on temporal discrimination in concurrent and post-feedback measures. *Perceptual and Motor Skills, 49*, 935-939.

Catania, A.C. (1998). *Learning* (4^{th} ed). Upper Saddle River, NJ: Prentice Hall

Domjan, M. & Burkhard, B. (1986). *The principles of learning and behavior* (2^{nd} ed.). Pacific Grove, CA: Brooks/Cole.

Ford, J.E. (1980). A classification system for feedback procedures. *Journal of Organizational Behavior Management, 2*(3), 183-191.

Hemmes, N.S., Brown B.L. & Kladopoulos, C.N. (2004). Time perception with and without a concurrent nontemporal task. *Perception & Psychophysics, 66*(2), 328-341.

Hicks, R.E. & Miller, G.W. (1976). Transfer of time judgments as a function of feedback. *American Journal of Psychology, 89*, 303-310.

Kazdin, A.E. (1989). *Behavior modification in applied settings* (4^{th} ed.). Pacific Grove, CA: Brooks/Cole.

Kladopoulos, C.N., Brown, B.L., Hemmes, N.S. & Cabeza de Vaca, S. (1998). The start-stop procedure: Estimation of temporal intervals by human participants. *Perception and Psychophysics, 60*, 438-450.

Mallott, R.W. (2008). *Principles of behavior* (6^{th} ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Mangiapanello, K.A. & Hemmes, N.S. (2009). An analysis of feedback from a behavior-analytic perspective. Unpublished manuscript.

Montare, A. (1988). Further learning effects of knowledge of results upon time estimation. *Perceptual and Motor Skills, 66*, 579-588.

Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst, 5*, 101-102.

Prue, D.M & Fairbank, J.A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management, 3*(1), 1-16.

Rakitin, B.C. (2005). The effects of spatial stimulus-response compatibility on choice time production accuracy and variability. *Journal of Experimental Psychology: Human Perception & Performance*, *31*(4), 685-702.

Robinson, E. (1963). The effect of degree of knowledge of results on time estimation. *The Psychological Record, 13*, 355-360.

Stolarz-Fantino, S. Fantino, E. & Van Borst, N. (2006). Use of base rates and case cue information in making likelihood decision. *Memory & Cognition, 34*(3), 603-618.

Stevens, S.S. (1960). The psychophysics of sensory function. *American Scientist, 48,* 226-253.

Sulzer-Azaroff, B. & Mayer, G.R. (1991). *Behavior analysis for lasting change*. Orlando, FL: Harcourt Brace Jovanovich.