Prospective and retrospective duration judgments: A meta-analytic review

RICHARD A. BLOCK
Montana State University, Bozeman, Montana

and

DAN ZAKAY
Tel Aviv University, Tel Aviv, Israel

A meta-analytic review compared prospective and retrospective judgments of duration, or duration judgment paradigm. Some theorists have concluded that the two paradigms involve similar cognitive processes, whereas others have found that they involve different processes. A review of 20 experiments revealed that prospective judgments are longer and less variable than are retrospective judgments. Several theoretically important variables moderate these effects, especially those concerned with information processing activities. Therefore, somewhat different cognitive processes subserve experienced and remembered duration. Attentional models are needed to explain prospective judgments, and memory-based models are needed to explain retrospective judgments. These findings clarify models of human duration judgment and suggest directions for future research. Evidence on duration judgments may also influence models of attention and memory.

Timing is essential to ensure the optimal functioning of organisms. People may look both backward and forward into time, remembering past events and anticipating future events. Under many conditions, people focus mainly on present events. A person must encode temporal properties of important events, construct cognitive representations of those properties, and use those representations for actions. For example, driving a car requires a person to estimate durations in order to engage in appropriate actions at the correct time.

Perhaps because the time dimension plays such an important role, the psychology of time was one of the important topics of early psychological research and theorizing (Block, 1990; Michon & Jackson, 1985). We may view human temporal experiences in several ways, depending on whether we emphasize simultaneity, successiveness, temporal order, duration, or temporal perspective. We focus here on duration, especially in the range of seconds and minutes. Duration timing in this range is essential for representing the present and recently past external environment. The experience of duration is the most

heavily researched aspect of psychological time, possibly because it is the most complex and important aspect in terms of environmental adaptation. It is puzzling that no single sensory organ or perceptual system subserves psychological time. This state of affairs has led most theorists to explain duration experiences in terms of cognitive processes or interactions between cognitive and biological processes (such as involving internal clocks). Because duration timing requires attention and memory, duration judgments reveal and clarify those cognitive processes (Block & Zakay, 1996).

James (1890) made a crucial distinction between different kinds of duration experiences: He proposed that different variables influence the "retrospective and prospective sense of time" (p. 624). Researchers have investigated this claim in two kinds of experiments. One involves asking participants prospectively to expose stimuli at a rate of one per subjective second until stopped by the experimenter, then asking them retrospectively to estimate the total time period (e.g., Frankenhaeuser, 1959; Hicks, 1992). The typical finding is that the retrospective verbal (numerical) estimate is less than the total duration prospectively produced. However, this kind of experiment does not afford a legitimate comparison of prospective and retrospective timing, because the duration is judged with two different methods.

A more common way to test James's distinction is by comparing duration estimates in what researchers now call the *prospective paradigm* and the *retrospective paradigm*. In the prospective paradigm, participants know in advance that they will be asked to judge the duration of a time period. In the retrospective paradigm, participants do not know until after a time period that they are being asked to judge its duration. In both cases, of course, par-

ticipants experience a time period in passing. The way in which they experience it and the various cognitive processes involved may nevertheless differ between the two paradigms. In the prospective paradigm, a person may intentionally encode temporal information as an integral part of the experience of the time period. This is partly why Block (1990) and others have used the term *experienced duration* to refer to the prospective paradigm. In the retrospective paradigm, a person may incidentally encode temporal information, and whatever information is relevant may be retrieved from memory later. Hence, the term *remembered duration* refers to the retrospective paradigm.

Many studies have used a prospective paradigm, but relatively few have used a retrospective paradigm. The main reason for this imbalance is that, after a participant is asked to provide a retrospective judgment, the participant is then aware that he or she may be asked to judge a subsequent duration. This is the defining characteristic of prospective judgment.[1] Some studies have directly compared duration judgments under the two paradigms, and these are the present focus. Gilliland, Hofeld, and Eckstrand (1946) questioned whether the duration judgment paradigm would influence duration judgments, but their review did not report any evidence. To our knowledge, Bakan (1955) conducted the first experiment. However, only a few studies reviewed here cited his experiment, perhaps partly because he found no significant difference between prospective and retrospective duration judgments.

Investigations of the duration judgment paradigm did not become common until after Hicks, Miller, and Kinsbourne's (1976) seminal, widely cited study. Hicks et al. asked participants to sort playing cards according to a rule that required processing zero bits, one bit, or two bits of information per card. In the prospective paradigm, subsequent verbal estimates of the 42-sec duration were an inverse linear function of amount of information processed. In the retrospective paradigm, duration judgments were unaffected by the amount of information processed. More convincing evidence that prospective and retrospective duration judgments involve somewhat different processes comes from a few experimental findings suggesting that several variables differentially influence judgments in the two paradigms. For example, Block (1992) replicated Hicks et al.'s finding: Experienced duration decreased when a processing task was more difficult, but remembered duration was not significantly affected. In a second experiment, Block found that remembered duration increased when participants performed several different kinds of tasks during the duration, but experienced duration was not significantly affected. These findings, revealing a double dissociation, provide relatively strong evidence that different processes or systems subserve the two kinds of duration judgments.

As a result of such evidence, some theorists have emphasized differences in processes subserving prospective and retrospective judgments (e.g., Block, 1992; Hicks et al., 1976). Most theorists think that experienced dura-

tion increases when a person allocates more attentional resources to processing temporal information. According to attentional models (e.g., Thomas & Weaver, 1975; Zakay & Block, 1996), a person divides attentional resources between nontemporal (stimulus) and temporal information. Thus, experienced duration should increase if the number of stimuli requiring processing is small, if a processing task is easy, if participants do not need to actively respond to presented information, or if they do not need to divide attention between two sources of stimuli.

Most theorists think that retrospective duration judgments increase as a function of the amount of stored and retrieved information, or storage size (Ornstein, 1969), the number of remembered changes (Fraisse, 1957/1963), the number of encoded and available contextual changes (Block & Reed, 1978), or the degree of segmentation of events during the time period (Poynter, 1983). Thus, theorists typically propose memory-based models. Attention to time may play little or no role in remembered duration unless a person has little information to process, frequent feelings of boredom, and so on. Such conditions may arouse a temporal motive (Doob, 1971).

In spite of the evidence that prospective and retrospective judgments differ, some theorists have emphasized the essential similarity of the timing processes involved. For example, Brown (1985) concluded that "the most important feature of these results is the similarity of prospective and retrospective judgments" (p. 119). This statement is striking, especially because he found that prospective judgments were longer and more accurate than were retrospective judgments. Furthermore, Brown and his colleagues (Brown, 1985; Brown & Stubbs, 1988, 1992) have consistently found paradigm effects equal to or larger than the mean paradigm effect in the studies reviewed here.

Whether prospective duration judgments are inherently larger in magnitude than are retrospective judgments may depend partly on whatever variables influence the ability of participants to allocate attention differentially to temporal information or to stimulus information. Thus, the relative length of prospective and retrospective judgments may depend on several variables. Finding an overall difference in the magnitude of duration judgments in the two paradigms is consistent with the notion that different processes are involved. Of course, magnitude differences may simply reflect different rates of similar underlying processes. Finding that different variables selectively influence prospective and retrospective duration judgments would strongly suggest that different processes subserve the two kinds of judgments.

The question of whether prospective and retrospective judgments differ in intersubject variability has not received much attention in theorizing about effects of duration judgment paradigm. We expect greater intersubject variability when different people use different processes and these processes affect duration judgment magnitude differently. The special instructions given in the prospective paradigm probably lead those participants to attend to time, thereby providing at least one common process.

As a result, intersubject variability may be smaller in the prospective paradigm than in the retrospective paradigm.

We report here the results of a meta-analytic review of experiments on the duration judgment paradigm. The review first answers the question of whether prospective and retrospective duration judgments differ in magnitude. More importantly, variables that may moderate any obtained difference in duration judgment magnitude are investigated in order to clarify the processes involved in each paradigm. The finding of significant moderator variables is vital information that sheds light on proposed models. (In a meta-analysis, the finding of a moderator variable is analogous to the finding of an interaction effect in primary research.) Because all reviewed studies reported the duration length used, we also accumulated primary-level statistics on a common scale of measurement: the ratio of subjective to objective duration. In this regard, the present review differs from most meta-analytic reviews: Comparisons of primary statistics clarify the meta-analytic statistics. We also review evidence on the intersubject variability of duration judgments made under the two paradigms. Finally, evidence is considered concerning the slope of the psychophysical function under the two paradigms.

METHOD

Sample of Studies

We searched a database containing about 9,000 references on the psychology of time (Block & H. Eisler, 1996). It includes articles from the following sources: *Psychological Abstracts* (1923–1966), *PsycINFO* (1967–1972), and *PsycLIT* (1973–1996), using the keywords *time perception* and *time estimation*; and *Medline* (1966–1996), using the keyword *time perception*.[2] The database also includes references from published bibliographies on time research (Das, 1990; H. Eisler et al., 1980; Krudy, Bacon, & Turner, 1976; Macey, 1991; Zelkind & Sprug, 1974), references from numerous articles, book chapters, and books (e.g., Doob, 1971; Fraisse, 1957/1963), and references from our files. For the present meta-analyses, we also searched *Social Sciences Citation Index* (Social SciSearch, 1977–1996) for articles that cited Hicks et al. (1976). Finally, we checked the reference lists of all relevant studies.

To be included in the present meta-analyses, an experiment must have involved normal human participants judging durations predominantly greater than 5 sec. These criteria would have excluded any experiment using very short durations, animals, participants showing gross psychopathology, or participants experiencing an unusual physical condition or an altered state of consciousness; however, there were none. Therefore, the only effective criterion for including an experiment is that it manipulated duration judgment paradigm as a between-subjects variable (i.e., each participant made only one duration judgment) and used a duration judgment method that allowed us to compare the magnitude, variability, or psychophysical function of judgments in the two para-

digms. We excluded experiments in which participants made more than one duration judgment, thereby confounding duration judgment paradigm and order of presentation—that is, experiments in which paradigm was a within-subjects variable, with the retrospective condition run first (by necessity). These included studies by Dobson (1954) and Unrug-Neervoort, Kaiser, Coenen, and van Luijtelaar (1991). In them, a positive time-order effect, in which the first of two equal durations is judged longer, may have obscured any paradigm effect (see Block, 1985). We also had to exclude two experiments by Predebon (1996) in which participants judged the relative duration of time periods with varying content.

Coded Variables

Mainly on the basis of a consideration of theoretically relevant variables, we coded the following variables from each experiment and from each within-experiment condition: (1) publication year, (2) temporal motive aroused by an unusually boring or repetitive task in the retrospective paradigm (no, probably, or yes), (3) participants' watch removal (no, yes, unspecified, or only from participants in prospective condition), (4) number of stimuli (none, one, several, or many), (5) stimulus complexity (simple, moderate, complex, or not applicable), (6) duration length (short [5.0–14.9 sec], moderate [15.0–59.9 sec], or long [60.0 sec or longer]), (7) processing type (passive/covert [passive perception of stimuli], active/covert [active processing, such as counting items of a certain type, without overt responding to each stimulus], or active/overt [active processing with overt responding to each stimulus]), (8) memory type (incidental [task instructions did not require memory for presented information] or intentional [task instructions required memory for presented information]), (9) processing difficulty or level[3] (easy/structural, moderate, difficult/semantic, or not applicable), (10) attentional demands (easy/unitary [e.g., attending to one stimulus/message only], moderate/selective [e.g., attending to one of two stimuli/messages], difficult/divided [e.g., simultaneously attending to two or more stimuli/messages], or not applicable), and (11) duration judgment method (verbal estimation, reproduction, comparison, repeated production, or analogical/absolute). The authors coded all study attributes independently, resolving any disagreements by discussion.

We also coded several other variables, but, for these, only one class of the variable was adequately represented in the present meta-analyses; the others contained fewer than three effect size estimates. No conclusions may be drawn for such variables. Future research is needed to reveal whether these variables moderate the paradigm effect on duration judgments. Nearly all experiments (1) studied participants of both sexes or did not report participants' sex, (2) studied young adults, usually college students, (3) did not manipulate participants' body temperature, (4) did not use any arousal manipulation, (5) used visual stimuli, (6) used moderately or highly familiar stimuli, (7) did not have any salient markers or high-priority

events segmenting the time period, (8) had few environmental or background changes, (9) had no changes in type or level of processing, and (10) used no delay, except possibly for brief instructions, preceding the duration judgment.

Effect Size Analyses

The authors independently estimated effect sizes, resolving disagreements by discussion. Each effect size was calculated as g, the difference between the mean duration judgment given by participants in each paradigm divided by the pooled standard deviation (Hedges & Olkin, 1985), using the computer software DSTAT (Johnson, 1989, 1993). Effect sizes were calculated separately, whenever possible, for different levels of manipulated variables (e.g., for different duration lengths). To provide a single measure for each experiment, we averaged all such separately calculated effect sizes. Each g was then converted to a d by correcting it for bias, weighting by the reciprocal of its variance, TW (Hedges, 1981; Hedges & Olkin, 1985). This procedure gives greater weight to effect sizes that are more reliably estimated. Then, the ds were combined by calculating unweighted and weighted means.

If a study manipulated a potential moderator variable and provided adequate information to calculate separate effect size estimates for each level of the variable, we did so for that moderator analysis. Thus, each moderator analysis contained a mixture of study effect sizes and within-study effect sizes. Using more than one effect size estimate from the same study violates the assumption that the effect sizes are independent. However, this kind of violation does not substantially affect statistical precision (Tracz, 1984/1985; Tracz, Elmore, & Pohlmann, 1992). If we had not used more than one effect size estimate from experiments that manipulated a potential moderator variable, we could not have properly conducted the moderator analyses, because we would have had to discard (or code into a mixed category) some of the most relevant information.

The homogeneity of each set of ds was tested to determine whether the conditions shared a common effect size. If there was significant heterogeneity of effect sizes, as indicated by the statistic Q, we attempted to account for it with coded or manipulated study attributes. One coded variable, publication year, is continuous. This variable was fit with a weighted least squares regression model (Hedges, 1982b; Hedges & Olkin, 1985), using SPSS and DSTAT (see Johnson, 1989). The regression model yields a significance test of the predictor variable, Q_R, as well as a test of the completeness or fit of the model, Q_E, which indicates whether significant unexplained variance remained in the effect sizes. All other coded attributes are categorical. For these, we used categorical models (Hedges, 1982a; Hedges & Olkin, 1985), as implemented by DSTAT. (In a few cases, we combined two similar classes of a variable when there were fewer than three effect size estimates in a given class.) These techniques yield a between-classes effect, which

reveals whether that variable is a significant moderator of the paradigm effect (i.e., whether one would expect to find an interaction effect in a primary experiment). It is estimated by Q_B, which is distributed approximately as X^2 with $k - 1$ degrees of freedom, where k is the number of classes. The homogeneity of effect sizes within each class was estimated by Q_{wi}, which has an approximate X^2 distribution with $m - 1$ degrees of freedom, where m is the number of effect sizes in the category. The mean effect size (d_{i+}) was calculated for each category, with each effect size weighted by the reciprocal of its variance in order to give greater weight to effect sizes that were estimated more reliably. If Q_B was significant and there were more than two classes of the moderator variable, we used DSTAT to perform post hoc paired contrasts of weighted mean effect sizes (d_{i+}) in each pair of classes. All p values for these simple contrasts are two-tailed.

For each meta-analysis, the order in which coded attributes (potential moderator variables) are listed reflects our judgment about the relative importance (from most to least) of each variable. This judgment was based on several criteria: (1) the size and significance of the between-classes effect in the relevant categorical model, (2) the completeness of the categorical model as indicated by within-class heterogeneity of variance, and (3) differences in the primary-level statistic (see next paragraph). We also calculated correlations among coded variables in order to determine the extent to which they were relatively independent of each other. We subsequently classified potential moderators into three categories: variables that influence only prospective duration judgments, those that influence only retrospective duration judgments, and those that do not differentially influence duration judgments in the two paradigms.

Primary-Level Statistics

For each condition analyzed, we calculated the ratio of subjective to objective duration (hereafter called the *duration judgment ratio*) separately for each paradigm. Researchers fairly routinely calculate and report this measure (Hornstein & Rotter, 1969), including authors of some studies reviewed here (e.g., Brown, 1985). This ratio enables a comparison of duration judgments across conditions and experiments that used different durations. The mean ratio of prospective/retrospective judgments (hereafter called the *paradigm ratio*) was also calculated. Accumulating these primary-level statistics across conditions and experiments clarifies the meta-analytic statistics (Pierce & Block, 1996).[4] Two-tailed independent-samples t tests were performed on these unweighted primary-level statistics. Although we do not report the t values, the description of results takes into account whether or not any particular comparison was significant (at $p < .05$).

We also analyzed separately the experiments that provided sufficient information, such as standard deviations or standard errors, to determine the relative intersubject variability of prospective and retrospective judgments.

If the two kinds of judgments differ in magnitude, one cannot simply compare standard deviations, because they typically increase with increasing mean judgment when a ratio scale of measurement is involved. We instead used the common psychometric measure, the coefficient of variation, which is the standard deviation divided by the mean judgment. Because no experiment reviewed here reported an analysis of coefficient of variation data, the present meta-analysis reveals new information about intersubject variability of duration judgments in the two paradigms. The program COEFVAR (Gilpin, 1993) was used to calculate a X^2 value for the difference between coefficients of variation with the Bennett–Shafer–Sullivan likelihood-ratio test (Shafer & Sullivan, 1986). DSTAT was then used to convert each X^2 to d. Primary-level statistics on coefficient of variation were also accumulated and tested.

Finally, we evaluated studies that reported data concerning the slope, or exponent (β), of the psychophysical function relating subjective to objective duration. Differences in the exponent reveal whether paradigm interacts with duration length.

RESULTS

Characteristics of Experiments

A total of 20 experiments, published in 16 separate references (14 journal articles and 2 book chapters), met all criteria for inclusion in at least one meta-analysis.[5] A total of 15 references were written in English, and 1 appeared originally in Spanish. The median publication year was 1985.

Duration Judgments

Effect size and primary statistics. Each of 18 experiments contributed an effect size for duration judgments in the two paradigms (see Appendix A). A total of 12 were calculated from an F or t value for paradigm, 2 from means and a related F value, and 4 from means and standard deviations or standard errors. The sign of each effect size was positive when the duration judgment ratio (the ratio of subjective to objective duration) was greater in the prospective paradigm and was negative when it was greater in the retrospective paradigm. The resulting weighted mean effect size $d_+ = 0.35$, 95% confidence interval (CI) = 0.26–0.43, indicating a significantly greater duration judgment ratio for prospective judgments than for retrospective judgments ($p < .0001$). The homogeneity statistic indicated that effect sizes were not homogenous [$Q(17) = 222.8, p < .0001$], and coded study attributes were used to account for variability in effect sizes.

The d_+ of 0.35, although significant, is considered small in magnitude (Cohen, 1977). The mean duration judgment ratio was significantly greater in the prospective paradigm than in the retrospective paradigm, and the mean paradigm ratio was greater than 1. Even though the effect size was small, the paradigm ratio revealed a rela-

tively large difference: The mean duration judgment ratio was 16% greater for prospective judgments than for retrospective judgments. This finding is consistent with the notion that the two kinds of duration judgments may involve somewhat different processes. However, this hypothesis clearly also depends on the finding of moderator variables.

Moderator variables. Table 1 shows the results of categorical model testing involving theoretically important variables that were sufficiently represented across experiments or frequently manipulated in experiments. It displays the between-classes effect (Q_B), the number of effect sizes (k) and the weighted mean effect size for each class (d_{i+}), the 95% CI for each effect size, and the homogeneity statistic for each class (Q_{wi}).

Both the meta-analytic and the primary-level statistics suggest that several variables differentially influence duration judgments in the two paradigms. We first discuss one variable that influences only prospective duration judgments, then two variables that influence only retrospective duration judgments, and then several variables that either do not influence either kind of duration judgment or do not differentially influence judgments in the two paradigms. In the Summary and Theoretical Discussion section, we consider the impact of the present findings on duration judgment theories. Finally, we suggest some directions for future research.

One variable influenced only prospective duration judgments. One variable, processing difficulty, selectively influenced only prospective duration judgments.

Processing difficulty. The paradigm effect was significant only for easy processing: It was larger for easy processing than for moderate processing ($p < .001$) and for difficult processing ($p < .01$), but the latter did not differ significantly. Because the easy and moderate classes showed heterogeneity of variance, processing difficulty was not the only moderator of the paradigm effect. Prospective duration judgment ratios decreased as processing difficulty increased (0.94, 0.82, and 0.73, respectively, for easy, moderate, and difficult processing), but retrospective ratios did not (0.78, 0.76, and 0.67, respectively). As a result, the paradigm ratio decreased as processing difficulty increased (1.20, 1.08, and 1.07, respectively). Processing difficulty was therefore an important moderator variable, affecting only prospective judgments.

Difficulty of information processing during the target duration is probably the most important moderator variable revealed by the present meta-analysis. As difficulty increases, experienced duration decreases. This finding is consistent with the notions that a person may divide attention between nontemporal and temporal information and that attention to time is a major process underlying experienced duration. As processing difficulty increases, opportunities to attend to time decrease. Attention to time in the prospective paradigm may be responsible for the finding that prospective judgments are usually larger in magnitude than are retrospective judgments. Attention

Table 1
Tests of Categorical Models for Duration Judgment Comparisons

Variable and Class	Between-Classes Effect (Q_B)	k	Mean Effect Size (d_{i+})	95% CI for d_{i+} Lower	95% CI for d_{i+} Upper	Homogeneity Within Class $(Q_{wi})^a$
Processing difficulty	29.18‡					
Easy		16	0.56	0.45	0.68	123.58‡
Moderate		7	0.02	−0.15	0.20	94.77‡
Difficult		8	0.19	−0.01	0.38	7.71
Duration length	27.04‡					
Short		4	−0.13	−0.34	0.07	166.63‡
Moderate		5	0.54	0.39	0.69	3.53
Long		10	0.35	0.23	0.47	19.08*
Stimulus complexity	15.35‡					
Simple		15	0.45	0.35	0.54	111.56‡
Moderate or complex		4	0.02	−0.17	0.21	91.75‡
Watch removal	35.89‡					
No or unspecified		12	0.52	0.42	0.62	95.81‡
Yes		6	−0.04	−0.19	0.11	87.12‡
Processing type	5.12*					
Covert		12	0.26	0.15	0.36	125.73‡
Overt		9	0.46	0.32	0.59	86.88‡
Temporal motive	4.30*					
No		14	0.31	0.22	0.40	206.50‡
Probably		4	0.60	0.34	0.87	3.18
Memory type	2.30					
Incidental		15	0.30	0.21	0.40	202.13‡
Intentional		4	0.46	0.28	0.65	15.08†
Number of stimuli[b]	1.49					
One		6	0.26	0.10	0.42	209.41‡
Several or many		13	0.38	0.28	0.48	27.54*
Duration judgment method[b]	0.35					
Verbal estimation		8	0.51	0.39	0.64	6.89
Reproduction		9	0.46	0.34	0.59	102.07‡
Attentional allocation	0.23					
Unitary		18	0.36	0.27	0.45	253.58‡
Selective or divided		5	0.45	0.10	0.79	0.54

Note—Effect sizes are more positive for a greater duration judgment ratio in the prospective than in the retrospective paradigm. [a]Significance indicates rejection of the hypothesis of homogeneity. [b]The category mixed or unknown was excluded from this analysis. *$p < .05$. †$p < .01$. ‡$p < .001$.

to time does not seem to be intimately involved in the retrospective paradigm, because processing difficulty did not affect remembered duration.

Two variables influenced only retrospective duration judgments. Two variables, duration length and stimulus complexity, selectively influenced only retrospective duration judgments.

Duration length. Conditions that used short durations showed no significant paradigm effect, whereas those that used moderate and long durations did. Effect sizes were smaller for short durations than for moderate and long durations (both *p*s < .001), and the latter did not differ significantly. Because the short and long classes showed significant heterogeneity of variance, duration length was not the only moderator of the paradigm effect. Prospective duration judgment ratios were comparable for short durations (0.90) and moderate or long durations (0.92). Retrospective ratios, however, were greater for short durations (0.90) than for moderate or long durations (0.78). As a' result, the paradigm ratio was smaller in conditions using short durations (1.02) than in those using moderate or long durations (1.17). Duration length was therefore an important moderator variable, affecting

only retrospective judgments. Other evidence consistent with this finding came from studies comparing the slope of the psychophysical function relating judged duration to actual duration (see later).

The finding that duration length influences remembered duration is consistent with some, but not all, memory-based models. It does not seem consistent with a storage size hypothesis (Ornstein, 1969), which has no reason to predict that encoded stimulus events require less storage space as duration lengthens. However, a contextual change model (Block & Reed, 1978) predicts this finding, because the amount of contextual change per unit of time is assumed to decrease as a person continues to engage in a certain type of activity. Most models of experienced duration, such as attentional models, do not assume that the amount of attention a person allocates to time decreases as the duration lengthens; if anything, they may predict a slight effect in the opposite direction. As a result, prospective duration judgments tend to be more accurate than retrospective judgments in that they usually are a more linear function of duration.

Stimulus complexity. Because only one effect size involved the presentation of moderately complex stimuli,

we combined it with three that involved the presentation of complex stimuli. Only conditions using simple stimuli showed a significant paradigm effect. Because both classes showed significant heterogeneity of variance, stimulus complexity was not the only moderator of the paradigm effect. Prospective duration judgment ratios were not significantly different for complex stimuli (1.01) and simple stimuli (0.87). However, retrospective ratios were greater for complex stimuli (0.94) than for simple stimuli (0.75). Stimulus complexity was therefore an important moderator, significantly affecting only retrospective judgments.

The finding that greater stimulus complexity led to increased remembered duration may be explained in terms of storage size or contextual change hypotheses. According to the latter, a person may encode a greater number of different interpretations of a more complex stimulus, leading to more changes in processing context (Block, 1990). However, findings on this issue are inconsistent. Stimulus complexity may influence remembered duration only if a person must actively process the information (Block & Zakay, 1996). The finding that stimulus complexity tends not to influence experienced duration is surprising. According to attentional models, a person should be able to allocate more attention to time if simple stimuli, rather than complex stimuli, are processed. Some recent evidence supports this prediction (Block & Zakay, 1996).

Several variables did not differentially influence duration judgments in the two paradigms. Several variables did not influence duration judgments in either paradigm or, if they did, did not differentially influence prospective and retrospective duration judgments.

Watch removal. Watch removal moderated effect sizes. Conditions in which participants' watches were removed showed no significant paradigm effect. The paradigm effect was greater in conditions that did not involve (or did not mention) watch removal than in those that did ($p <$ .0001). However, prospective duration judgment ratios were comparable for no watch removal (0.90) and watch removal (0.95), and retrospective ratios were also comparable for no watch removal (0.76) and watch removal (0.86). Thus, the paradigm ratio was comparable for experiments involving no (or unspecified) watch removal (1.19) and for those involving watch removal (1.11). These findings suggest that, although watch removal was a significant moderator, it was not an important one.

Some researchers have implicitly assumed that watch removal does not influence duration judgments. The present data show that experiments in which participants' watches were removed (usually with a cover story about jewelry interfering with the recording equipment) found a reduced (statistically eliminated) paradigm effect. This finding suggests that watch removal may influence participants to attend to time in the retrospective paradigm to nearly the same extent that they do in the prospective paradigm. We are not aware of any single study that has manipulated watch removal; such research could clarify this issue.

Processing type. Because only two effect sizes involved active/covert processing, we combined them with those that involved passive/covert processing. Processing type moderated effect sizes. Although the paradigm effect was greater when the primary task performed during the duration demanded responding to presented stimuli than when it did not ($p =$ .03), both classes showed a significant paradigm effect. Because both classes showed significant heterogeneity of variance, processing type was not the only moderator of the paradigm effect. Prospective duration judgment ratios were comparable for no-responding (0.94) and active-responding (0.86) conditions. Retrospective ratios were also comparable for no-responding (0.81) and active-responding (0.73) conditions. Because the paradigm ratio was similar for both classes (1.18 for no-responding conditions, and 1.19 for active-responding conditions), processing type was not an important moderator variable.

Although the categorical model revealed a significant influence of processing type on duration judgments, the duration judgment ratios were not significantly different between the two conditions. Attentional models of experienced duration predict that overtly responding to presented information should demand attention; therefore, the paradigm effect should decrease. Duration judgment ratios showed that this did not occur, although no single experiment in the present meta-analysis directly tested the effect of processing type. Predebon (1996) found somewhat inconsistent evidence on this issue: In Experiment 1, processing type did not influence either prospective or retrospective duration judgments; in Experiment 2, overt processing decreased prospective judgments and did not influence retrospective judgments. Future research should more thoroughly explore the conditions under which overt responding to presented information requires attention and thereby shortens experienced duration.

Publication year. The linear moderating influence of publication year was significant ($Q_R = 9.19, p = .002$), but the regression model did not provide a good fit ($Q_E = 212.3, p < .0001$). The scatterplot relating publication year and effect size revealed a negative correlation. However, removing two outliers (Zakay, 1992, 1993) eliminated the moderating influence of publication year ($Q_R = 1.47, p = .23$), and the regression model provided a reasonable fit ($Q_E = 22.5, p = .07$).

Temporal motive. Temporal motive is a coded variable that represented our assessment of the extent to which boring or repetitive conditions may have led participants in the retrospective paradigm to attend to time during the target duration (thereby making it more like the prospective condition). Evidence against this comes from the finding that the paradigm effect was marginally greater in experiments that probably aroused a temporal motive than in those that did not ($p = .04$). Prospective duration judgment ratios were comparable for experiments that did not arouse a temporal motive (0.89) and those that probably did (1.02). Retrospective ratios were also comparable for experiments that did not arouse a temporal motive (0.77) and those that probably did (0.88). Thus,

the paradigm ratio was similar for experiments that did not arouse a temporal motive (1.16) and those that probably did (1.17). In short, although temporal motive was a significant moderator, it was not an important one.

Memory type. Memory type did not moderate effect sizes ($p = .55$). Prospective duration judgment ratios were comparable for no-memory (0.91) and memory (0.90) conditions, and retrospective ratios were also comparable for no-memory (0.80) and memory (0.76) conditions. Thus, the paradigm ratio was comparable for memory (1.14) and no-memory (1.21) conditions.

These findings are surprising. We expected that processing tasks specifying memory for presented information (intentional memory tasks) would decrease experienced duration relative to those not specifying memory for presented information (incidental memory tasks). In the former, the attentional resources needed to attend to time should be reduced. However, only one experiment compared intentional and incidental memory conditions: McClain (1983) found that experienced duration was shorter when the task required intentional processing at a semantic level than when the instructions did not say that memory would be tested. Block and Zakay (1996) also reported that intentional memory instructions decrease experienced duration. Intentional memory instructions may indeed engender the allocation of additional attentional resources, although this question clearly warrants additional research.

Number of stimuli. Because only 2 effect sizes involved the presentation of several stimuli during the experimental duration, we combined those with 11 effect sizes that involved the presentation of many stimuli. Number of stimuli did not moderate effect sizes ($p = .22$). Prospective duration judgment ratios were comparable for one-stimulus (0.97) and several- or many-stimuli (0.90) conditions. Retrospective ratios were also comparable for one-stimulus (0.83) and several- or many-stimuli (0.78) conditions. Thus, the paradigm ratio was comparable for one-stimulus (1.18) and several- or many-stimuli (1.16) conditions.

These findings are also surprising. Some researchers (e.g., McClain, 1983; Predebon, 1996) have found that presenting more stimuli during a time period decreases experienced duration but increases remembered duration. Thus, the paradigm effect should have been greater under conditions in which fewer stimuli occurred. Other variables (e.g., processing difficulty or processing type) may be more important than the number of stimuli. Thus, the number of stimuli presented per se may not necessarily differentially influence experienced and remembered duration. Predebon (1996) found that prospective judgments decreased with greater stimulus quantity when overt responding to the stimuli was required, but not necessarily when only covert processing was required. He also found that retrospective judgments increased with greater stimulus quantity regardless of whether or not overt responding to the stimuli was required.

Duration judgment method. Only one effect size was computed from use of a comparison method and one from use of an analogical method (both by Zakay, 1993), so we eliminated those categories from further consideration. Conditions using verbal estimation and reproduction methods did not differ significantly ($p = .55$), suggesting that the paradigm effect was robust across those two methods. Prospective judgment ratios were comparable for verbal estimation (0.97) and reproduction (0.87), and retrospective ratios were also comparable for verbal estimation (0.80) and reproduction (0.76). Thus, the paradigm ratio was comparable for verbal estimation (1.23) and reproduction (1.14). There is little theoretical reason and no empirical evidence to suggest that verbal estimation and reproduction methods are differentially sensitive to paradigm differences.

Attentional allocation. Because only one effect size involved a condition in which participants selectively attended to one of two sources of stimuli, we combined it with four effect sizes that involved dividing attention between two sources. Attentional allocation did not moderate effect sizes ($p = .63$). Prospective duration judgment ratios were comparable for unitary (0.91) and selective or divided attention (0.80), and retrospective ratios were also comparable for unitary (0.77) and selective or divided attention (0.71). Thus, the paradigm ratio was comparable for unitary (1.20) and selective or divided attention (1.14).

Two experiments that directly manipulated attentional allocation (Brown, 1985, Experiment 2; Bueno Martínez, 1990) did not find a significant main effect or interaction of attentional allocation on prospective or retrospective duration judgments. This is surprising, because attentional models predict that experienced duration should decrease under divided-attention conditions. Along with our present finding, these findings also challenge an attentional selectivity model of remembered duration (Underwood, 1975; Underwood & Swain, 1973).

Coefficients of Variation

Effect size and primary statistics. Each of 14 experiments provided sufficient information (e.g., means and standard deviations) to calculate separate coefficients of variation for prospective and retrospective judgments (see Appendix B). The sign of each effect size was positive when the coefficient of variation was greater for retrospective judgments and negative when it was greater for prospective judgments. The resulting $d_+ = 0.11$, 95% $CI = 0.01-0.21$, indicating a greater coefficient of variation for retrospective than prospective judgments ($p = .03$). The homogeneity statistic indicated that effect sizes were not homogenous [$Q(13) = 39.4, p < .001$], and coded study attributes were used to account for variability in effect sizes.

The d_+ of 0.11, although significant, is considered small in magnitude. The mean coefficient of variation was significantly greater in the retrospective paradigm than in the prospective paradigm, and the mean paradigm ratio was greater than 1. Even though the effect size was small, the paradigm ratio revealed a relatively large difference: The mean coefficient of variation was 15%

greater for retrospective judgments than for prospective judgments.

Moderator variables. Table 2 shows the results of categorical model testing involving variables that both were sufficiently represented across experiments or frequently manipulated in experiments and were found to be significant moderator variables.

Processing difficulty. Processing difficulty moderated coefficient of variation effect sizes. The paradigm effect was larger for easy and difficult processing than for moderate processing (both $ps < .01$), and the former two classes did not differ significantly. Because the easy class showed heterogeneity of variance, processing difficulty was not the only moderator of the paradigm effect. For prospective judgments, coefficient of variation was smaller for easy processing (0.27) than for moderate or difficult processing (0.44). For retrospective judgments, coefficient of variation was comparable for easy processing (0.37) and moderate or difficult processing (0.43). The paradigm ratio was greater for easy processing (1.37) than for moderate or difficult processing (0.98). Processing difficulty was therefore an important moderator, affecting only prospective judgments.

The coefficient of variation analyses revealed that as processing difficulty increases, intersubject variability of prospective judgments also tends to increase, whereas variability of retrospective judgments is unaffected. The effect on variability of prospective judgments may reflect greater individual differences in the ability to attend to time as primary-task difficulty increases.

Duration length. Duration length also moderated effect sizes. Effect sizes were smaller for short durations than for moderate and long durations (both $ps < .01$), and the latter did not differ significantly. Conditions that used short durations showed no significant paradigm effect, whereas those that used moderate and long durations did. Because there was no significant heterogeneity of variance in the short and moderate classes and only relatively low (albeit significant) heterogeneity in the long

class, duration length provided a good account of the paradigm effect. In both paradigms, the coefficient of variation was smaller for short durations than for moderate or long durations, although not quite significantly so. In the prospective paradigm, the coefficients of variation were 0.19, 0.38, and 0.31, respectively, for short, moderate, and long durations. In the retrospective paradigm, they were 0.22, 0.49, and 0.36, respectively. The paradigm ratio was not significantly different between short (1.09), moderate (1.26), and long (1.20) durations.

Duration length may influence intersubject variability to a greater extent for remembered duration than for experienced duration. It is important to note that the coefficient of variation compensates for the similar moderating influence found in duration judgment ratios. As duration increases, different processes may subserve duration judgments in different participants, especially for remembered duration.

Attentional allocation. Attentional allocation also moderated effect sizes. The paradigm effect was greater under unitary attention conditions than under selective or divided conditions. Prospective coefficients of variation were comparable for unitary (0.29) and selective or divided attention (0.38), and retrospective coefficients of variation were also comparable for unitary (0.36) and selective or divided attention (0.40). The paradigm ratio, however, was marginally greater for unitary (1.26) than for selective or divided attention (0.91).

Thus, prospective judgment variability decreased under conditions in which participants did not have to selectively attend to or divide attention between two sources of information.

Other variables. The linear moderating influence of publication year on coefficient of variation was not significant ($Q_R = 2.11, p = .15$). There was also no moderating influence of watch removal, temporal motive, number of stimuli, stimulus complexity, processing type, processing changes, and duration judgment method (all $Q_Bs < 2.95$).

Table 2
Tests of Categorical Models for Coefficient of Variation Comparisons

Variable and Class	Between-Classes Effect (Q_B)	k	Mean Effect Size (d_{i+})	95% CI for d_{i+}		Homogeneity Within Class (Q_{wi})[a]
				Lower	Upper	
Processing difficulty	18.01‡					
Easy		12	0.22	0.10	0.35	27.55†
Moderate		5	−0.27	−0.47	−0.08	9.83
Difficult		5	0.21	−0.05	0.47	2.24
Duration length	11.67†					
Short		3	−0.19	−0.40	0.01	5.48
Moderate		4	0.25	0.07	0.43	3.97
Long		8	0.18	0.03	0.32	19.55*
Attentional allocation	4.27*					
Unitary		14	0.15	0.04	0.25	42.81‡
Selective or divided		5	−0.23	−0.57	0.11	8.69

Note—Effect sizes are more positive for a greater coefficient of variation in the retrospective than in the prospective paradigm.   [a]Significance indicates rejection of the hypothesis of homogeneity.   *$p < .05$.   †$p < .01$.   ‡$p < .001$.

Psychophysical Slope

For four experiments, published in three separate articles, the primary data analyses concerned the slope of the psychophysical function relating judged duration to actual duration (see Appendix C). Two effect sizes were computed from reported F values, one from reported means and the pooled standard deviation (calculated from a reported F value), and one by combining two separate p values (assuming the minimal effect size consistent with each reported value). The sign of each effect size was positive when the slope was greater for prospective judgments than for retrospective judgments. The resulting $d_+ = 0.69$, 95% $CI = 0.56$–0.82, indicating a greater slope for prospective judgments than for retrospective judgments ($p < .001$). The homogeneity statistic indicated that the effect sizes were not homogenous [$Q(3) = 19.7$, $p < .001$]. However, because the number of experiments was small, we did not use coded study attributes to account for variability in the effect sizes.

The mean d of 0.69 is considered moderate to large in magnitude. For both paradigms, the mean slope was less than 1. The mean slope was greater in the prospective paradigm (0.74) than in the retrospective paradigm (0.39). The mean prospective/retrospective slope ratio was 1.93, indicating that the prospective slope was nearly twice as large as the retrospective slope.

The finding that duration length moderated the paradigm effect on duration judgments is reflected here in the greater slope of the psychophysical function for experienced duration than for remembered duration. Whatever memory information underlies remembered duration, such as encoding of changes in cognitive context, may be encoded more frequently during the start of the duration than during an extended length of the duration. As duration lengthens, remembered duration does not increase proportionally.

SUMMARY AND THEORETICAL DISCUSSION

The present findings show that duration judgment paradigm influences mean duration judgment: Prospective judgments are longer than retrospective judgments. The weighted mean effect size (0.35), although small, has an associated confidence interval that does not include 0. Although both prospective and retrospective duration judgments tend to be underestimates, the overall paradigm ratio indicates that prospective judgments are about 16% greater than retrospective judgments. (Thus, prospective judgments are typically more accurate than retrospective judgments.) In addition, retrospective judgments show about 15% greater intersubject variability than do prospective judgments. These findings are consistent with the notion that somewhat different processes subserve experienced and remembered duration. As noted earlier, this hypothesis depends heavily on the finding of moderator effects (similar to interaction effects in a study reporting primary statistics). We indeed

found that several variables differentially influence experienced duration and remembered duration.

Earlier, we summarized canonical views on the processes involved in prospective and retrospective duration judgments. According to these views, prospective judgments depend heavily on attention to time. Increased attention to time in the prospective paradigm is presumably responsible for the major finding that prospective judgments are usually longer than retrospective judgments. A major finding here is that processing difficulty influences experienced duration in a way consistent with this idea: Prospective judgments are shorter when more of a person's attentional resources must be allocated to stimulus information processing instead of temporal information processing. This finding supports an attentional model of experienced duration. The evidence concerning several other variables that should influence attentional resources, however, was either inconclusive or negative. There was only weak evidence that requiring participants to respond to presented information also decreases experienced duration. There was no evidence that requiring participants to remember presented information has an influence or that requiring participants to attend to more than one source of presented information has an influence. Perhaps neither manipulation increased attentional demands, although that is a post hoc explanation. Additional research is needed on these and other related issues, involving both the collection of additional primary data and meta-analyses investigating experiments that manipulated these variables (e.g., Block & Zakay, 1996; Predebon, 1996).

As we noted earlier, no widespread agreement exists on how to explain remembered duration. The finding that retrospective judgments show greater intersubject variability than do prospective judgments suggests that participants may use more varied processes to judge remembered duration. The finding that processing difficulty does not influence remembered duration is consistent with the view that attention to time ordinarily plays little or no role in remembered duration. This may be qualified by the finding that watch removal may lengthen such judgments, perhaps by sensitizing people to the duration length and the resulting more frequent attending to time that they may display. Our finding that stimulus complexity influences remembered duration is consistent with a memory storage explanation (e.g., Ornstein's, 1969, storage size hypothesis), although the finding that the number of stimuli does not influence remembered duration weakens such an explanation. Several theorists have proposed that remembered duration increases when there are changes, such as in cognitive context (Block & Reed, 1978) or segmentation of presented information (Poynter, 1983).[6] Memory change models can handle the finding that retrospective duration judgments do not increase proportionally with duration length. They assume that more changes occur earlier in the time period than later, reflecting a process similar to that in positive time-order effects.[7] Although the effect of duration length is

consistent with memory-based explanations of remembered duration, the finding that several memory-related variables (e.g., intentional vs. incidental memory encoding) did not influence retrospective judgments was somewhat surprising. This may be attributable to the limited set of experiments considered in the present review. A better understanding of remembered duration may result from meta-analyses investigating experiments that manipulated these variables.

SUGGESTIONS FOR FUTURE RESEARCH

Prospective and retrospective duration judgments clearly differ in magnitude and variability. The present meta-analyses also reveal several variables that differentially influence experienced and remembered duration. The heterogeneity within each class of the moderator variables that produced significant effect size differences implies multiple interactions among variables. Of course, additional studies using a single duration judgment paradigm have also manipulated these variables. Other meta-analyses, focusing on effects of these other theoretically important variables, may reveal whether duration judgment paradigm is a significant moderator of those effects. These additional analyses may reveal the nature of the interactions more clearly. Only this broader set of meta-analyses, along with additional experimental evidence, will suggest more definitive conclusions regarding experienced and remembered duration.

Useful future experiments will explore the influences of theoretically relevant moderator variables. In particular, researchers should focus on clarifying effects and interactions involving the number of stimuli, processing type (covert vs. overt responding), and memory type (incidental vs. intentional). Additional studies investigating the psychophysical function will be important if they reveal variables that may interact with duration judgment paradigm.

Researchers who are not primarily interested in duration judgments per se should find some interesting implications of this work. For example, attention to time may compete with attention to stimulus information, especially if a task arouses a temporal motive. Performance of repetitive tasks may suffer for this reason alone. Remembered duration may reflect contextual changes during a time period, although this is somewhat less clear than the attentional effects on experienced duration. In the near future, studies of experienced and remembered duration may clarify general attention and memory processes (Block & Zakay, 1996; Zakay & Block, 1996).

REFERENCES

(References marked with an asterisk [*] indicate studies included in at least one meta-analysis.)

*Bakan, P. (1955). Effect of set and work speed on time estimation. *Perceptual & Motor Skills*, 5, 147-148.

Block, R. A. (1985). Contextual coding in memory: Studies of remembered duration. In J. A. Michon & J. L. Jackson (Eds.), *Time, mind, and behavior* (pp. 169-178). Berlin: Springer-Verlag.

Block, R. A. (1990). Models of psychological time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1-35). Hillsdale, NJ: Erlbaum.

*Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 141-152). Dordrecht: Kluwer.

Block, R. A., & Eisler, H. (1996). *The compleat bibliography on the psychology of time: 1839 to 1996* [Electronic data file]. Bozeman, MT: Montana State University, Department of Psychology (Producer and distributor).

*Block, R. A., George, E. J., & Reed, M. A. (1980). A watched pot sometimes boils: A study of duration experience. *Acta Psychologica*, 46, 81-94.

Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 656-665.

Block, R. A., & Zakay, D. (1996, October-November). *Duration judgments reveal attention and memory processes*. Poster presented at the annual meeting of the Psychonomic Society, Chicago.

Boltz, M. [G.] (1992a). The incidental learning and remembering of event durations. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 153-163). Dordrecht: Kluwer.

Boltz, M. G. (1992b). The remembering of auditory event durations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 938-956.

*Brown, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Perception & Psychophysics*, 38, 115-124.

*Brown, S. W., & Stubbs, D. A. (1988). The psychophysics of retrospective and prospective duration timing. *Perception*, 17, 297-310.

*Brown, S. W., & Stubbs, D. A. (1992). Attention and interference in prospective and retrospective timing. *Perception*, 21, 545-557.

*Bueno Martínez, M. B. (1990). Efectos de los cambios cognitivos y del esfuerzo de procesamiento sobre los juicios de duración prospectivos y retrospectivos. [Effects of cognitive changes and processing effort on judgment of prospective and retrospective duration]. *Estudios de Psicología*, 45, 109-127.

Bueno Martínez, M. B. (1992). Testing models of time estimation. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 173-176). Dordrecht: Kluwer.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Das, T. K. (1990). *The time dimension: An interdisciplinary guide*. New York: Praeger.

Dobson, W. R. (1954). An investigation of various factors involved in time perception as manifested by different nosological groups. *Journal of General Psychology*, 50, 277-298.

Doob, L. W. (1971). *Patterning of time*. New Haven, CT: Yale University Press.

Eisler, H., Linde, L., Troëng, G., Lazar, R., Eisler, B.-M., & Hellström, Å. (1980). A complementary bibliography of the psychology of time. *JSAS Catalog of Selected Documents in Psychology*, 10, 55 (Ms. No. 2101).

Fraisse, P. (1963). *The psychology of time* (J. Leith, Trans.). New York: Harper & Row. (Original work published 1957)

Frankenhaeuser, M. (1959). *Estimation of time: An experimental study*. Stockholm: Almqvist & Wiksell.

Gilliland, A. R., Hofeld, J. B., & Eckstrand, G. (1946). Studies in time perception. *Psychological Bulletin*, 43, 162-176.

Gilpin, A. R. (1993). A program to test equality of two or more coefficients of variation. *Behavior Research Methods, Instruments, & Computers*, 25, 65-66.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.

HEDGES, L. V. (1982a). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7,* 119-137.

HEDGES, L. V. (1982b). Fitting continuous models to effect size data. *Journal of Educational Statistics, 7,* 245-270.

HEDGES, L. V., & OLKIN, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

HICKS, R. E. (1992). Prospective and retrospective judgments of time: A neurobehavioral analysis. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 97-108). Dordrecht: Kluwer.

*HICKS, R. E., MILLER, G. W., & KINSBOURNE, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology, 89,* 719-730.

HORNSTEIN, A. D., & ROTTER, G. S. (1969). Research methodology in temporal perception. *Journal of Experimental Psychology, 79,* 561-564.

JAMES, W. (1890). *The principles of psychology.* New York: Henry Holt.

JOHNSON, B. T. (1989). *DSTAT: Software for the meta-analytic review of research literatures* [Documentation and computer program]. Hillsdale, NJ: Erlbaum.

JOHNSON, B. T. (1993). *DSTAT 1.10: Software for the meta-analytic review of research literatures* [Upgrade documentation and computer program]. Hillsdale, NJ: Erlbaum.

*KIKKAWA, M. (1983). The effect of distribution of attention upon prospective and retrospective estimations of long temporal intervals. *Journal of Child Development, 19,* 14-22.

*KINSBOURNE, M., & HICKS, R. E. (1991). The extended present: Evidence from time estimation by amnesics and normals. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 319-330). Cambridge: Cambridge University Press.

KRUDY, E. S., BACON, B. T., & TURNER, R. (1976). *Time: A bibliography.* London: Information Retrieval.

MACEY, S. L. (1991). *Time: A bibliographic guide.* New York: Garland.

*MCCLAIN, L. (1983). Interval estimation: Effect of processing demands on prospective and retrospective reports. *Perception & Psychophysics, 34,* 185-189.

MICHON, J. A., & JACKSON, J. L. (1985). Introduction: The psychology of time. In J. A. Michon & J. L. Jackson (Eds.), *Time, mind, and behavior* (pp. 2-17). Berlin: Springer-Verlag.

*MILLER, G. W., HICKS, R. E., & WILLETTE, M. (1978). Effects of concurrent verbal rehearsal and temporal set upon judgments of temporal duration. *Acta Psychologica, 42,* 173-179.

ORNSTEIN, R. E. (1969). *On the experience of time.* Harmondsworth, U.K.: Penguin.

PIERCE, C. A., & BLOCK, R. A. (1996). *Using primary-level statistics to improve meta-analytic interpretations.* Manuscript in preparation.

POYNTER, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition, 11,* 77-82.

*PREDEBON, J. (1995). Prospective and retrospective time estimates as a function of clock duration. *Perceptual & Motor Skills, 80,* 941-942.

PREDEBON, J. (1996). The effects of active and passive processing of interval events on prospective and retrospective time estimates. *Acta Psychologica, 94,* 41-58.

SHAFER, N. J., & SULLIVAN, J. A. (1986). A simulation study of a test for the equality of coefficients of variation. *Communications in Statistics: Simulation & Computation, 15,* 681-695.

THOMAS, E. A. C., & WEAVER, W. B. (1975). Cognitive processing and time perception. *Perception & Psychophysics, 17,* 363-367.

TRACZ, S. M. (1985). The effect of the violation of the assumption of independence when combining correlation coefficients in a meta-

analysis (Doctoral dissertation, Southern Illinois University, 1984). *Dissertation Abstracts International, 46,* 688A.

TRACZ, S. M., ELMORE, P. B., & POHLMANN, J. T. (1992). Correlational meta-analysis: Independent and nonindependent cases. *Educational & Psychological Measurement, 52,* 879-888.

UNDERWOOD, G. (1975). Attention and the perception of duration during encoding and retrieval. *Perception, 4,* 291-296.

UNDERWOOD, G., & SWAIN, R. A. (1973). Selectivity of attention and the perception of duration. *Perception, 2,* 101-105.

UNRUG-NEERVOORT, A., KAISER, J., COENEN, A., & VAN LUIJTE-LAAR, G. (1991). Prospective and retrospective time estimation. *Perceptual & Motor Skills, 73,* 1157-1158.

ZAKAY, D. (1989). Subjective time and attentional resource allocation: An integrated model of time estimation. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 365-397). Amsterdam: North-Holland.

*ZAKAY, D. (1992). The role of attention in children's time perception. *Journal of Experimental Child Psychology, 54,* 355-371.

*ZAKAY, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics, 54,* 656-664.

ZAKAY, D., & BLOCK, R. A. (1996). The role of attention in time estimation processes. In M. A. Pastor & J. Artieda (Eds.), *Time, internal clocks and movement* (pp. 143-164). Amsterdam: Elsevier, North-Holland.

*ZAKAY, D., & FALLACH, E. (1984). Immediate and remote time estimation—A comparison. *Acta Psychologica, 57,* 69-81.

ZELKIND, I., & SPRUG, J. (1974). *Time research: 1172 studies.* Metuchen, NJ: Scarecrow Press.

NOTES

1. An exception to this is illustrated by Boltz's (1992a, 1992b) research, in which participants were presented several events and retrospectively judged the duration of each.

2. We included any relevant study listed in *PsycLIT* or *Medline* as of November 1996.

3. For experiments manipulating attention (i.e., Block, George, & Reed, 1980; Brown, 1985, Experiment 2; Brown & Stubbs, 1992; Bueno Martinez, 1990), we coded processing difficulty independently of attentional demands by considering only the primary task or the information processed.

4. We also calculated weighted mean ratios, weighting each duration judgment ratio (and the paradigm ratio) by its sample size. This is similar to the weighting (by TW) that is involved in using ds as effect size estimates in the meta-analyses. Although doing so often increased the magnitude of the paradigm differences, it did not alter any conclusions. We report unweighted mean ratios and results of t tests based on them.

5. All 20 experiments contributed to the analyses of duration judgment ratios. Three of the experiments were also separately reported in additional publications; see the footnotes accompanying Appendixes A, B, and C for details.

6. Bueno Martinez (1990) found no effect of processing changes in either paradigm. Block (1992, Experiment 2) found that processing changes increase remembered duration but do not influence experienced duration. In addition, Block and Zakay (1996) reported that processing changes increase remembered duration and decrease experienced duration.

7. When participants judge the first of two equal durations as being longer than the second, the data show a positive time-order effect. Cognitive context may ordinarily change more rapidly at the start of a new experience, such as being a participant in an experiment (see Block, 1985).

APPENDIX A
Duration Judgment Ratios and Effect Sizes

Study	Pro Ratio	Pro n	Ret Ratio	Ret n	Pro/Ret Ratio	Total N	Effect Size (d)
Bakan (1955)	1.00	22	0.90	22	1.11	44	0.24
Hicks, Miller, & Kinsbourne (1976)	1.00	60	0.77	60	1.31	120	0.34
Miller, Hicks, & Willette (1978)	0.89	126	0.59	126	1.51	252	0.72
Block, George, & Reed (1980, Exp. 1)[a]	0.96	24	0.87	24	1.10	48	0.46
Block, George, & Reed (1980, Exp. 2)[a]	1.07	24	0.85	24	1.26	48	0.96
Block, George, & Reed (1980, Exp. 3)[a]	1.06	48	0.91	48	1.17	96	0.68
Kikkawa (1983)	1.17	36	1.08	36	1.08	72	0.22
McClain (1983)[a]	0.81*	90	0.77*	90	1.05*	180	0.14*
Zakay & Fallach (1984, Exp. 3)[b]	0.87	40	0.72	40	1.20	80	0.40
Brown (1985, Exp. 1)[a]	0.94	48	0.81	48	1.16	96	0.53
Brown (1985, Exp. 2)[a]	0.83	48	0.63	48	1.32	96	0.52
Brown & Stubbs (1988, Exp. 1)	1.03*	72	0.87*	86	1.18*	158	0.66*c
Bueno Martínez (1990)[d]	0.43	96	0.41	96	1.06	192	0.09*c
Kinsbourne & Hicks (1991, Exp. 2)[e]	0.93	151*	0.73	151*	1.27	302	—
Block (1992, Exp. 1)	0.75	60	0.62	60	1.22	120	0.55
Block (1992, Exp. 2)	0.66	60	0.65	60	1.03	120	0.08
Brown & Stubbs (1992)	1.16*	96	1.03*	79	1.13*	175	—
Zakay (1992)[a]	0.95	32	0.77	80	1.23	112	2.93
Zakay (1993)	0.80	112	1.00	112	0.79	224	−1.11
Predebon (1995)	0.98	108	0.86	108	1.15	216	0.43
Overall mean (unweighted)[f]	0.91*	20	0.79*	20	1.16*	20	0.49*
Overall mean (weighted)[g]	0.89*	1353*	0.77*	1398*	1.16*	2751	0.35*

Note—Positive effect size (d) indicates that the duration judgment ratio (subjective/objective duration) was greater for the prospective paradigm; negative effect size (d) indicates that it was greater for the retrospective paradigm. A dash (—) indicates that the authors did not provide sufficient data for us to estimate an effect size (see slope analysis). Pro = prospective; Ret = retrospective. [a]Data from second duration judgment (repeated measure) were excluded. [b]These data were also reported in Zakay (1989, Study 1). [c]Overall d was estimated from means shown in a figure or table and a related F value. [d]These data were also reported in Bueno Martínez (1992). [e]These data were also reported in Hicks (1992, Table 1). [f]Each mean weights each experiment equally. [g]Each mean weights by n contributing to each datum (or by TW for effect size). *An approximate datum (e.g., one estimated from a figure).

APPENDIX B
Coefficients of Variation (CV) and Effect Sizes

Study	Pro CV	Pro n	Ret CV	Ret n	Ret/Pro CV	Total N	Effect Size (d)
Bakan (1955)	0.37	22	0.48	22	1.30	44	0.31
Hicks, Miller, & Kinsbourne (1976)	0.57	60	0.82	60	1.44	120	0.35
Block, George, & Reed (1980, Exp. 1)[a]	0.14	24	0.26	24	1.86	48	0.91
Block, George, & Reed (1980, Exp. 2)[a]	0.25	24	0.20	24	0.80	48	−0.26
Block, George, & Reed (1980, Exp. 3)[a]	0.21	48	0.23	48	1.10	96	0.12
Kikkawa (1983)	0.34	36	0.28	36	0.82	72	−0.24
Brown (1985, Exp. 1)[a]	0.28	48	0.41	48	1.46	96	0.49
Brown (1985, Exp. 2)[a]	0.34	48	0.32	48	0.94	96	−0.06
Bueno Martínez (1990)[b]	0.55	96	0.64	96	1.16	192	0.16
Block (1992, Exp. 1)	0.28	60	0.46	60	1.46	120	0.64
Block (1992, Exp. 2)	0.35	60	0.32	60	0.91	120	−0.08
Zakay (1992)[a]	0.08	32	0.07	80	0.88	112	−0.20
Zakay (1993)	0.22	112	0.18	112	0.82	224	−0.31
Predebon (1995)	0.34	108	0.41	108	1.21	216	0.24
Overall mean (unweighted)[c]	0.31	14	0.36	14	1.15	14	0.15
Overall mean (weighted)[d]	0.33	778	0.37	826	1.12	1604	0.11

Note—Positive effect size (d) indicates that the coefficient of variation was larger for the retrospective paradigm; negative effect size (d) indicates that it was larger for the prospective paradigm. Pro = prospective; Ret = retrospective. [a]Data from second duration judgment (repeated measure) were excluded. [b]These data were also reported in Bueno Martínez (1992). [c]Each mean weights each experiment equally. [d]Each mean weights by n contributing to each datum (or by TW for effect size).

APPENDIX C
Psychophysical Slopes and Effect Sizes

Study	Pro		Ret		Pro/Ret	Total	Effect
	Slope	n	Slope	n	Slope	N	Size (d)
Brown & Stubbs (1988, Exp. 1)[a]	0.63	72	0.32	86	1.97	158	1.07
Brown & Stubbs (1988, Exp. 2)[a]	0.75	179	0.38	171	1.97	350	0.91
Kinsbourne & Hicks (1991, Exp. 2)[b]	0.87*	151*	0.48*	151*	1.81*	302*	0.31*[c]
Brown & Stubbs (1992)	0.60	96	0.30	79	2.00	175	0.66
Overall mean (unweighted)[d]	0.71*	4	0.3 /*	4	1.94	4	0.67*
Overall mean (weighted)[e]	0.74*	498*	0.39*	487*	1.93*	985*	0.69*

Note—Positive effect size (d) indicates that the slope was greater for the prospective than the retrospective paradigm. Pro = prospective; Ret = retrospective. [a] Data from the second duration judgment (repeated measure) are excluded. [b] These data were also reported in Hicks (1992, Table 1). [c] Overall d was estimated by combining two inexact p values. [d] Each mean weights each experiment equally. [e] Each mean weights by n contributing to each datum (or by TW for effect size). *An approximate datum (e.g., one estimated from a figure).