Memory & Cognition
2000, 28 (8), 1333-1346

Sex differences in duration judgments: A meta-analytic review

RICHARD A. BLOCK
Montana State University, Bozeman, Montana

PETER A. HANCOCK
University of Minnesota, Minneapolis, Minnesota

and

DAN ZAKAY
Tel Aviv University, Tel Aviv, Israel

We quantitatively reviewed human sex differences in the magnitude and variability of duration judgments. Data from 4,794 females and 4,688 males yielded 87 effect size estimates of magnitude and 28 of variability. The overall sex difference in duration judgment magnitude was small but statistically significant. It was moderated by whether study participants knew in advance (prospective paradigm) or only later (retrospective paradigm) that they would be required to judge duration. Although prospective judgments showed no overall sex effect, some levels of moderator variables showed a small but statistically significant effect. Retrospective judgments showed a larger subjective-to-objective duration ratio for females than for males, and several variables moderated this effect. Females' judgments also showed more intersubject variability than did males' judgments. Relative to males, females sustain attention to time more in the prospective paradigm and have better episodic memory in the retrospective paradigm.

Since the publication of a landmark study on human sex differences in various abilities and behaviors (Maccoby & Jacklin, 1974), researchers have become increasingly interested in the topic. Researchers have conducted more than 100 meta-analyses on human sex differences. These differences range from nonexistent (effect size d not significantly different from 0) to small ($d = 0.2$ or less) to large ($d = 0.8$ or greater). For example, the sex difference on spatial perception tasks is small, but that on mental rotation tasks is large (Linn & Petersen, 1985).

It is difficult to detect a general pattern in the evidence on sex differences involving memory and cognitive processes. Females perform relatively better on tasks involving production and comprehension of complex prose, fine motor skills, or perceptual speed, whereas males perform better on tasks involving visuospatial transformations, spatiotemporal operations, or fluid reasoning (Halpern, 1997). Females also perform better on some memory tasks than do males (Halpern, 2000). Males and females do not differ on working memory and semantic memory tasks, but females perform relatively better on episodic memory tasks, including recall and recognition of words, facts, and performed activities, than do males (Herlitz, Airaksinen, & Nordstroem, 1999; Herlitz, Nilsson, & Bäckman, 1997). Although the female advantage on episodic memory tasks is small ($d = 0.05$ to 0.34), it has been found in at least 18 experiments, whereas a male advantage has been found in only 2 experiments (Herlitz et al., 1997).

Our review focuses on a relatively unexplored topic that has direct implications for understanding sex differences in memory and cognitive processes: sex differences in psychological time. Some researchers have found such sex differences, although the issue is far from settled. Psychological time involves processes by which a person adapts to and represents temporal properties in order to synchronize actions with external events. For example, while driving down or crossing a busy street, speed and time estimates are continually required. While waiting in a line or at a Web site, feelings of lengthened duration may influence whether a person completes the transaction and returns for a subsequent visit. A person may use a duration judgment to decide whether to continue trying to solve a problem or to quit. Because many everyday perceptual and cognitive situations lead a person to estimate short durations, it is important to understand the underlying processes and whether there are individual differences in them.

This review was supported in part by grants from the United States–Israel Binational Science Foundation and the MONTS Program at Montana State University. A preliminary version of this research was presented at the meeting of the International Society for the Study of Time, July 1995. The authors thank Robin Hall for retrieving some articles, Iren Putashkin, Dida Sasson, and Fusako Matsui for translating some articles, and Erik Arthur for coding some moderators. The authors also thank Scott Brown, Alice Eagly, Hannes Eisler, Diane Halpern, Charles Pierce, and two anonymous reviewers for helpful comments on a preliminary version of this article. Correspondence should be addressed to R. A. Block, Department of Psychology, Montana State University, Bozeman, MT 59717-3440 (e-mail: block@montana.edu).

SEX DIFFERENCES IN PSYCHOLOGICAL TIME

Psychological time has been a topic of research and theorizing for more than a century (Block & Zakay, in press), probably because the temporal dimension plays such a crucial role in human life. In early essays, several theorists focused on duration judgments. Some of the first research was conducted in the context of psychophysical investigations (Woodrow, 1951). More recently, researchers study duration judgment processes to clarify more general processes involving attention and memory. Our research pursues this focus.

Research on sex-related differences in duration judgments began a century ago (MacDougall, 1904; Seashore, 1899; Yerkes & Urban, 1906). Early researchers thought that females made relatively larger and more variable estimates of duration than did males, although not all investigators found this pattern (e.g., Swift & McGeoch, 1925). Researchers offered very few explanations. By the mid-1960s, sex differences were rarely mentioned in duration judgment research, and researchers who did suitable analyses usually reported finding no significant sex difference. Researchers who had conducted experiments on duration judgments with equal numbers of males and females often did not report separate statistics for the two sexes. Beginning in the late 1980s and early 1990s, some researchers again considered the issue. Some found a sex difference that seemed to echo what was revealed in early research (H. Eisler & A. D. Eisler, 1992; Hancock, Arthur, Chrysler, & Lee, 1994; Hancock, Vercruyssen, & Rodenburg, 1992). Others thought that the literature showed that females underestimate durations more than do males (Kellaris & Mantel, 1994). Explanations for these more recent findings remain elusive. One eminent reviewer concluded: "Sex differences [in estimates of duration] have not been reliably established" (Fraisse, 1998, p. 496). Another researcher (Hancock, 1999), who published an extensive qualitative review, concluded that there are sex differences in both the magnitude and the variability of duration judgments. He attributed these sex differences to differences in spatiotemporal perceptual capabilities. No researcher has published a quantitative review of sex differences in duration judgments.

MODELS OF DURATION JUDGMENT PROCESSES

Theorists differ on how to explain duration judgments (Block, 1990). One view emphasizes physiological processes. In the most common variant of this view, an internal clock consisting of a pacemaker and additional components subserves time-related behavior. Variables, such as brain temperature, metabolism, psychoactive drugs, and arousal level, may influence the pacemaker rate. Another view emphasizes that duration is a cognitive construction that is influenced mainly by processes involving attention and memory. Various views emphasize different variables, so choosing a suitable method to investigate sex differences in duration judgments is critical. Two methodological variables are particularly important: the duration judgment paradigm and the duration judgment method.

Duration Judgment Paradigm

Undoubtedly, the most important variable influencing duration judgment magnitude is the actual target duration. Arguably, the next most important variable influencing duration judgments is whether a person knows in advance that a duration estimate will be required. In the prospective paradigm, a person has this knowledge; in the retrospective paradigm, a person does not. The magnitude and variability of duration judgments crucially depend on this variable: Prospective judgments are larger in magnitude and smaller in variability than retrospective judgments, and different variables moderate judgments in the two paradigms (Block & Zakay, 1997).

A consensus is emerging regarding the influence of the duration judgment paradigm. In the prospective paradigm, duration judgment magnitude is directly related to the amount of attention allocated to temporal information processing during the target duration. Subjective duration decreases if a person must process more attention-demanding nontemporal information. In essence, prospective time estimation is a divided-attention task. If there is a sex difference in the ability to divide attention, we expect to find an interaction of sex and processing difficulty. However, the meager extant evidence reveals little sex difference in the ability to divide attention (e.g., Herlitz et al., 1997; Seth-Smith, Ashton, & McFarland, 1989). If this is the case, we expect to find little sex difference in the prospective paradigm.

In the retrospective paradigm, duration judgment magnitude is directly related to the amount of encoded and retrieved memory information, such as concerning events or contextual changes. Remembered duration decreases if events or changes are fewer or more difficult to remember (Block, 1978, 1990). In essence, retrospective time estimation is an episodic memory task. As noted earlier, research has revealed a female superiority on episodic memory tasks (Herlitz et al., 1999; Herlitz et al., 1997). Thus, we expect to find a sex difference in retrospective duration judgments, with females remembering durations as being relatively longer than do males.

Because duration judgment paradigm is such an important variable, it is fortunate that the literature contains sufficient experiments using each paradigm in order to assess its role as a moderator variable. This enables us to distinguish attention- and memory-based accounts of sex differences in duration judgments.

Duration Judgment Method

The method used to obtain duration judgments is another important variable. In the verbal estimation method, a person is asked to use conventional time units to subjectively estimate an experienced (objective, or target) duration. In the production method, a person is asked to delimit an objective duration corresponding to a subjectively defined target duration (e.g., "Hold this button down for what seems like 60 sec to you"). In a variant on this method, the repeated-production method, a person is not asked to produce a single duration; instead, he or she is asked to delimit consecutive durations of a specified length, usually 1 sec (e.g., "Press this button once every second until I tell you to stop"). Here, we combine data on these two production methods. Verbal estimation and production methods are similar in that they involve comparing a duration experience with information stored in memory concerning conventional duration units, such as seconds and minutes. They have drawbacks in some contexts because they assume that the translation between conventional units and subjective duration is reliable. They may be suitable, however, to investigate individual differences in use of conventional units (or effects of variables that may influence the rate of internal processes). Researchers have successfully used them in studies of factors thought to influence the rate of internal timekeeping processes, such as stimulant drugs (e.g., Frankenhaeuser, 1959; Hicks, 1992).

In the reproduction method, the person experiences a duration and then is asked to delimit a subjective duration to estimate it. The method is similar to verbal estimation in that it involves estimating an already-experienced duration and is similar to production in that it involves making an operative estimate. Although reproductions do not require a translation involving conventional duration units, the method has another potential drawback: If the rate of physiological and cognitive processes varies from one individual to another, the same rate subserves an individual's experiencing the target duration and reproducing the duration. Thus, the reproduction method may not reveal anything interesting concerning sex differences in duration judgments. It may only detect individual differences in the framework of psychophysical studies, in which duration is varied, or if some important variable (such as processing difficulty) is manipulated during the target duration but not during the reproduction. In addition, extraneous variables (e.g., a person's desire to terminate the experiment as quickly as possible or some more general form of impatience) may confound findings obtained using the production and the reproduction methods.

GOALS OF THE META-ANALYTIC REVIEW

Our goals were to evaluate evidence concerning sex differences in human duration judgments and to discover what variables moderate any differences. Moderator variables are those that account for variability among effect sizes across experimental conditions. The outcomes clarify theories concerning processes that may produce sex differences, such as those involving attention and memory. We separately evaluated sex differences in duration judgment magnitude and variability. Evidence of interindividual variability bears on the issue of whether participants of each sex use similar processes (see Morse, 1993). If participants use different processes (or if the intraindividual variability of similar processes is greater) and these processes affect duration judgments differently, we expect to find greater interindividual variability. On most cognitive measures, within-sex variability is greater for males than for females (Feingold, 1992; Hedges & Nowell, 1995). We also evaluated several studies of the slope of the psychophysical function relating subjective and objective duration, another potential index of sex differences in psychological time.

METHOD

Sample of Studies

We searched a database containing more than 10,000 references on the psychology of time (Block & Eisler, 2000). It includes articles from PsycINFO (1887–1999), using the keywords *time perception* and *time estimation*, and from Medline (1966–1999), using the keyword *time perception*.[1] It also includes other references (see Block & Zakay, 1997, for details). We searched *Social Sciences Citation Index* (Social SciSearch, 1977–1999) for articles that cited relevant articles.

To be included in our meta-analyses, experimenters must have studied normal humans, and they must have reported analyzing quantitative data on duration judgment magnitude separately for females and males. Thus, we excluded experiments that studied animals, humans showing gross psychopathology, or participants experiencing an unusual physical condition or an altered state of consciousness. We also excluded (1) articles that contained only absolute-error data, percentage underestimation and overestimation, and other such measures, (2) experiments involving other temporal judgments, such as duration discrimination, (3) experiments using noncontinuous durations, and (4) experiments using designs that confounded important variables (e.g., males and females judged different target durations).

In experiments that primarily investigated other variables but that studied participants of both sexes, some researchers did not report including sex as a factor in the data analyses. We did not include these experiments in our meta-analyses.

Effect Size Analyses

Each effect size was calculated as g, the difference between the mean duration judgment made by participants of the two sexes divided by the pooled standard deviation (Hedges & Olkin, 1985), using the computer program DSTAT (Johnson, 1989, 1993). Whenever possible, we calculated effect sizes separately for different levels of manipulated variables (e.g., for different target durations). To provide a single estimate for each experiment, we averaged all such separately calculated effect sizes. If an experimenter reported only a nonsignificant finding, with no inferential statistics, we estimated g by assuming that $p = .50$ if we could determine the direction of the effect from either duration judgment means or a verbal description or by assuming that $g = 0$ if we could not determine the direction of the effect. We converted each g to a d by correcting it for bias, which involved weighting it by the reciprocal of its variance (Hedges, 1981; Hedges & Olkin, 1985). This procedure gave more weight to effect sizes that were more reliably estimated. We then summarized the effect size estimates by separately calculating unweighted and weighted mean effect sizes.

Coded Variables and Moderator Analyses

We tested the homogeneity of each set of ds to determine whether the conditions shared a common effect size. We attempted to explain heterogeneity of effect sizes (as indicated by the Q statistic) in terms of coded or manipulated attributes. Continuous moderator variables were tested by using a weighted-least squares regression model (Hedges, 1982b; Hedges & Olkin, 1985), using SPSS and DSTAT (see Johnson, 1989). Categorical moderator variables were tested by using a categorical model (Hedges, 1982a; Hedges & Olkin, 1985), as implemented by DSTAT. (We combined two similar classes of a variable if there were fewer than three effect size estimates in a given class.) These techniques yield a between-classes effect, revealing whether that variable is a significant moderator of the sex effect. Finally, we assessed the relative importance of each significant moderator variable by relying on several statistics: (1) the results of an analysis in which significant moderator variables were entered into a multiple regression equation (Johnson, 1989), (2) the size and significance of the between-classes effect in the categorical model, and (3) the completeness of the categorical model as indicated by each within-class heterogeneity of variance (Johnson & Turco, 1992).

If a researcher manipulated a potential moderator variable and provided adequate information for us to calculate separate effect size estimates for each level of the variable, we did so for that moderator analysis. Thus, following Cooper's (1989) recommendation, each moderator analysis contained both experiment and within-experiment effect sizes.

After considering theoretically relevant variables, we decided to code the following potential moderator variables from each experiment and from each within-experiment condition:

1. *Publication year*. This was the copyright date of the article.

2. *Participants' age*. This was coded based on the mean age (or other such information) reported: *children* (8.0–12.9 years old); *adolescents* (13.0–17.9 years old); *young adults*, such as all studies involving college students (18.0–29.9 years old); *old adults* (30.0–59.9 years old), or *older adults* (at least 60.0 years old).

3. *Duration judgment paradigm*. This was coded as *prospective* (if the description of the procedure stated that participants were informed in advance that they would make a duration judgment, or if they made more than one judgment) or as *retrospective* (if the description did not state that participants were informed in advance and they made only one judgment).

4. *Number of stimuli*. This was coded as *none* (i.e., an "empty" duration, one not containing any presented stimulus), *one* (e.g., a continuous tone), *several* (e.g., three words), or *many* (e.g., a film or a passage of text).

5. *Modality of stimuli*. This was the dominant modality of any presented information: *visual* (e.g., words on a screen), *auditory* (e.g., a series of tones), *tactile*, or *mixed* (e.g., a film with sound and video).

6. *Complexity of stimuli*. This was coded as *simple* (e.g., nonmusical tones or common words), *moderate* (e.g., a passage from a novel), or *complex* (e.g., a passage of scientific text).

7. *Segmentation of stimuli*. This was coded as *none* (if there were no high-priority events, or discontinuities that segmented the duration; see Poynter, 1989), *moderate* (if there were a few such events), or *high* (if there were many such events).

8. *Duration length*. This was coded, on the basis of the predominant length of the target duration, as *very short* (less than 5.0 sec), *short* (5.0–14.9 sec), *moderate* (15.0–59.9 sec), or *long* (60.0 sec or longer).

9. *Processing type*. This was coded as *passive and covert* (e.g., listening to text), *active and covert* (e.g., counting items of a certain type, without overtly responding to each stimulus), or *active and overt* (e.g., executing a motor response to each stimulus).

10. *Processing difficulty*. This was coded as *easy* (e.g., listening to a tone), *moderate* (e.g., performing a Stroop test), or *difficult* (e.g., making semantic decisions about word categories).

11. *Duration judgment method*. This was coded as *verbal estimation*, *production*, *repeated production*, *reproduction*, or *other* (see earlier).

12. *Duration judgment immediacy*. This was coded as *immediate* (if a production was made or if a verbal estimation or reproduction was made within about a minute after the target duration ended) or *delayed* (if a verbal estimation or reproduction was made more than a minute after the target duration ended).

13. *Number of trials*. This was the total number of duration judgments made by each participant in the experiment (or in an experimental condition).

Two of us coded the experiment attributes independently, resolving disagreements by discussion. For some variables, we used the categories *mixed*, *unknown*, or *not applicable* (e.g., if no stimulus was presented, complexity of stimuli is not applicable). The moderator analyses excluded experimental conditions with such coding. We also coded several other variables (see Block & Zakay, 1997), but, for these, only one class of the variable was adequately represented; the others contained fewer than three effect size estimates. We can draw no conclusions about these variables.

Primary Statistics

Whenever separate data were available for each sex, we calculated what we call the *duration judgment ratio*—the ratio of subjective-to-objective duration—for each experimental condition. Many researchers report this measure (see Block, Zakay, & Hancock, 1998, 1999). For the verbal estimation method, this is the ratio of the person's numerical (subjective) estimate to the target (objective) duration. For the production method, this is the ratio of the requested (subjective) duration to the person's operative (objective) duration estimate. Production is the methodological inverse of verbal estimation (Bindra & Waksberg, 1956; Zakay, 1990), and this ratio reverses the typically negative correlation between estimates obtained using the two methods. Thus, the ratio assesses the moderating influence of duration judgment method apart from the otherwise negative correlation. For the reproduction method, the duration judgment ratio is the ratio of the person's reproduced (subjective) duration to the previously presented (objective) duration. We also calculated the ratio of female-to-male duration judgment ratios—hereafter called the *sex ratio*. These data were available from 50 of 87 (57%) of the experiments. We cautiously used these statistics to clarify the moderator variable analyses: Comparing duration judgment ratios across moderator variable levels reveals whether an effect is seen in females, males, or both. It also provides evidence on the accuracy of the mean duration judgments made by females and males. We performed correlated two-tailed t tests on unweighted duration judgment ratios.

We also analyzed separately the experimental conditions for which reported data (such as standard deviations or standard errors) allowed us to determine the interindividual variability of duration judgments made by female and male participants. Few researchers have commented on any sex difference in variability, so our meta-analysis reveals new information about interindividual variability in duration judgments.[2] On a ratio scale, standard deviations typically increase proportionally with judgment magnitude. To obviate this problem, we used a common psychometric measure called the *coefficient of variation*, which is the standard deviation divided by the mean judgment. The program COEFVAR (Gilpin, 1993) was used to calculate a chi-square value for the difference between coefficients of variation with the Bennett–Shafer–Sullivan likelihood ratio test (Shafer & Sullivan, 1986). We used DSTAT to convert each chi-square to an effect size. We accumulated primary statistics on coefficients of variation.

Finally, we evaluated experiments that investigated the slope of the psychophysical function relating subjective-to-objective duration across different duration lengths.

RESULTS AND DISCUSSION

A total of 87 experiments, published in 71 journal articles and 2 book chapters, met all criteria for inclusion in our review. A total of 68 articles were written in English, 3 in Japanese, 1 in Russian, and 1 in Spanish. The median publication year was 1980.

Duration Judgment Magnitude

Overall effect size and primary statistics. A total of 87 experiments, which entailed data from 4,794 females and 4,688 males, contributed an effect size estimate for the magnitude of duration judgments made by females compared with males (see the Appendix). We calculated 27 effect size estimates from means and standard deviations, standard errors, quartile deviations, or average deviations;[3] 15 from reported t or F values (including one estimated from related F values); 1 from a χ^2 value; 2 by assuming an exact value (e.g., $p = .001$) for a reported inexact p value (e.g., $p < .001$); 6 by assuming that $p = .5$ (effects reported as nonsignificant for which we were able to infer the direction of the effect); and 36 by assuming that $g = 0$ (effects reported as nonsignificant for which we were unable to infer the direction of any effect). We defined the sign of each effect as positive if the duration judgment ratio was larger for females and as negative if it was larger for males. The resulting weighted mean effect size [$d_+ = 0.03$, 95% confidence interval (CI) = 0.00 to 0.07] indicated a larger duration judgment ratio for females than males ($p = .05$). The unweighted mean effect size ($d = 0.16$) was greater than the weighted mean effect size. The homogeneity statistic revealed that effect sizes were heterogeneous [$Q(86) = 300.2$, $p < .0001$], so we used coded attributes to explain the variability in effect sizes.

Researchers who conducted 36 of the 87 experiments either (1) reported including sex as a factor in their data analysis but provided no descriptive or inferential statistics on it or (2) reported finding no significant sex difference but did not provide enough information to estimate an effect size. Because we included these studies, for which we assigned an effect size $d = 0$, the overall weighted mean effect size may have underestimated the actual effect size (Becker, 1996). Some researchers (e.g., Eagly, Ashmore, Makhijani, & Longo, 1991) have argued that meta-analysts should exclude experimental reports that do not contain sufficient information to compute at least one effect size. When we included only the 51 experiments for which sufficient statistics were reported, the weighted mean effect size ($d_+ = 0.06$, 95% CI = 0.01 to 0.11) indicated a larger duration judgment ratio for females than for males ($p = .02$). The unweighted mean effect size ($d = 0.27$) was greater than the weighted mean effect size. These findings are surprising in that they suggest that the experiments for which $d = 0$ were not substantially different from the other experiments.[4] However, one difference was in the sample size. Experiments that contained sufficient statistics had almost twice the mean sample size ($N = 135.2$) as those that did not ($N = 71.8$). Experiments that had less power to detect a significant effect tended not to include sufficient statistics on the potential sex effect.[5]

For the 49 experimental reports that provided primary statistics on duration judgment magnitude, the mean duration judgment ratio was larger for females than for males [$t(48) = 3.13$, $p = .003$]. The duration judgment ratio for females (1.21) was significantly greater than 1.00 [$t(48) = 2.01$, $p = .05$], whereas that for males (1.06) was not significantly different from 1.00 [$t(48) = 1.06$, $p = .30$]. The sex ratio (1.10) was significantly greater than 1.00 [$t(47) = 3.32$, $p = .002$]. In short, although the weighted mean effect size was small, the sex ratio in duration judgments revealed a fairly large difference: The mean duration judgment ratio was 10% greater for females than for males.

Overall moderator analyses. Nine variables significantly moderated the overall sex effect: participants' age, duration judgment paradigm, complexity of stimuli, duration length, environmental changes, processing difficulty, duration judgment method, duration judgment immediacy, and number of trials. However, some variables were probably significant only because the levels of them were correlated with the levels of other variables that actually moderated the sex effect. In particular, characteristics of experiments that were conducted using the prospective paradigm usually differed from characteristics of experiments that were conducted using the retrospective paradigm. We regard the duration judgment paradigm as the most important moderator variable, one that influenced many other moderators. The duration judgment paradigm showed a between-classes effect [$Q_B(1) = 6.77$, $p = .01$]. The sex effect did not differ from zero for conditions that used the prospective paradigm [$d_+ = 0.01$, 95% CI = −0.03 to 0.06], but it was positive for conditions that used the retrospective paradigm [$d_- = 0.16$, 95% CI = 0.06 to 0.26]. To reveal potentially different moderators of sex differences in each paradigm, we conducted separate meta-analyses for prospective and retrospective judgments.

Prospective paradigm. A total of 74 experiments contributed an effect size estimate for the prospective paradigm. Five variables were significant moderators: participants' age, processing difficulty, duration length, duration judgment method, and number of trials. For all five moderators, however, one or more classes showed significant heterogeneity of variance; thus, no single variable provided a complete model (Johnson & Turco, 1992). Several other coded variables were sufficiently represented across studies or frequently manipulated in experiments, but there was no significant moderation by publication year, number of stimuli, modality of stimuli, complexity of stimuli, segmentation of stimuli, processing type, or duration judgment immediacy.

A moderator variable may have been significant only because its levels were correlated with the levels of another significant moderator. In fact, duration length was correlated with participants' age [$r(25) = .37$, $p = .03$] and duration judgment method [$r(25) = .68$, $p < .001$]. Participants' age was correlated with duration judgment method [$r(25) = .60, p < .001$]. We entered the five significant moderator variables into a multiple regression model, with the criterion variable *d* weighted by the reciprocal of its variance. Duration judgment method, a variable that contained three classes, was dichotomously coded by using *1* for the method that did not require the use of conventional duration units (i.e., reproduction) and *2* for the methods that required the use of such units (i.e., verbal estimation and production). This model did not account for a significant amount of the variability in effect sizes. The best-fitting regression model contained only three predictors [$R = .56$, $Q_R(3) = 56.8, p < .001$]: number of trials (standardized regression coefficient $\beta = 0.11, p = .001$), participants' age ($\beta = -0.16, p = .001$), and duration judgment method ($\beta = 0.17, p = .02$). The fit was not as good when we added processing difficulty, duration length, or both, and neither variable was significant. Figure 1 shows results of model testing of the three significant moderators.

Number of trials. The mean number of trials was 23.5. Number of trials moderated effect sizes [$Q_R(1) = 39.2$, $p < .0001$], and the correlation between effect size and number of trials was large [$r(72) = .65, p < .001$]. Experiments in which participants made more duration judgments tended to find a more positive sex effect. There are at least two possible explanations, which are not mutually exclusive: (1) Considering that there is substantial intraindividual and interindividual variability in duration judgments (Doob, 1971), increasing the number of trials may have decreased the variability of the judgments made by participants of each sex, thereby revealing more clearly the veridical effect; and (2) as the number of trials increased, so did the total duration of the experiment. (Of course, no report contained data on the latter, so this is speculative.) Some sex-related factors may have become accentuated across the experimental session, such as decreasing likelihood of attending to time or increasing boredom. Support for this explanation comes from the fact that the correlation between the number of trials and the duration judgment ratio was not significant for females [$r(35) = -.15$], but it was for males [$r(35) = -.39$]; this difference between correlations was significant [$t(34) = 2.08, p < .05$]. The males' decreased ratio of subjective-to-objective duration as the number of trials increased may reflect a decreased probability that males attend to time during each target duration.

Participants' age. There was no significant difference in weighted mean effect size between adolescents and young adults ($p = .55$), nor between old adults and older adults ($p = .99$), so we combined those pairs of age

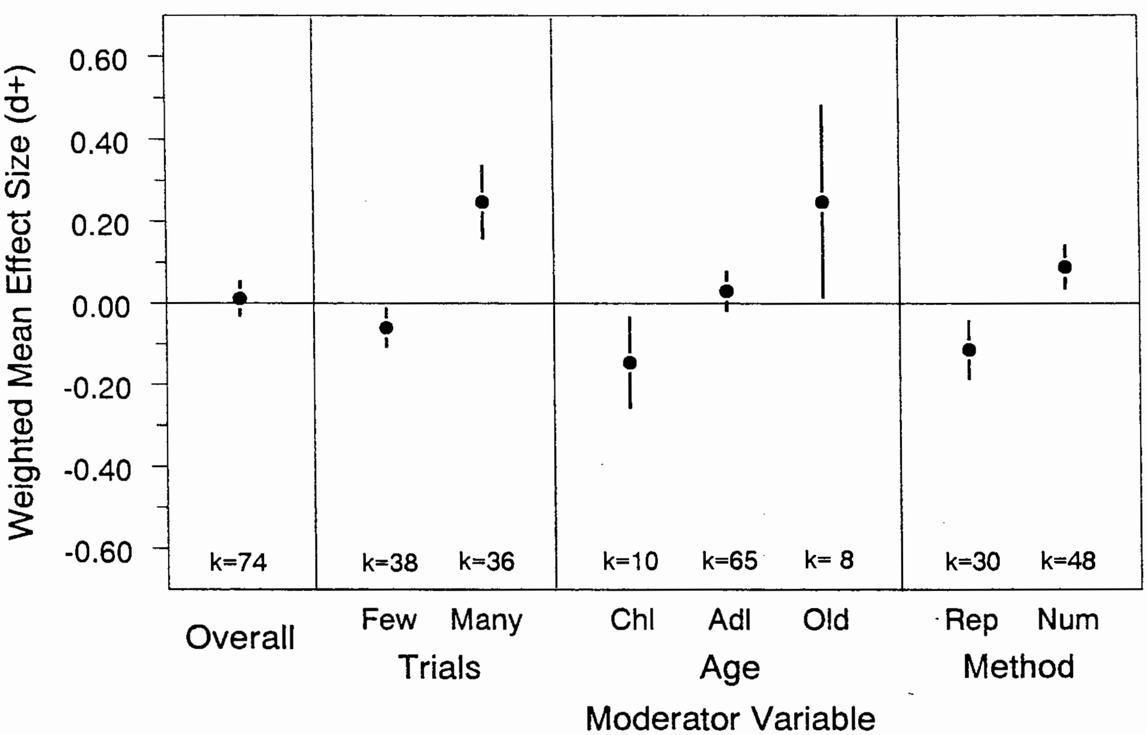

Figure 1. Weighted mean effect size (d_+) and 95% CI in the prospective paradigm (Overall) and for each class of a moderator variable that a multiple regression analysis revealed to be an important predictor. The number of experimental conditions (*k*) that contributed to each weighted mean is shown. Trials = number of trials (shown here coded according to a median split for comparison purposes, although it was analyzed as a continuous variable); Few = 0–10 trials; Many = 12–240 trials; Age = participants' age; Chl = children; Adl = adolescents and young adults; Old = old adults and older adults; Method = duration judgment method; Rep = reproduction; Num = method using numerical time units (i.e., verbal estimation or production).

classes. Participants' age showed a significant between-classes effect [$Q_B(2) = 21.8, p < .001$]. The sex effect was negative for children, not significant for adolescents and young adults, and positive for old adults and older adults.

Duration judgment method. Duration judgment method showed a between-classes effect [$Q_B(1) = 21.8, p < .001$]. The sex effect was positive for conditions that used the production or verbal estimation method, and it was negative for conditions that used the reproduction method. The effect in conditions that used the reproduction method differed from that in conditions that used the verbal estimation ($p = .003$) and production methods ($p < .001$). The effect size statistics reveal that, relative to males, females made shorter reproductions, and they made shorter productions and larger verbal estimates. The duration judgment ratios suggest that females (0.89) and males (0.89) underreproduce durations to the same extent. Females significantly underproduce durations (1.13), but males do not (1.00). Females verbally overestimate durations (1.10), but males do not (0.98). Thus, the main finding is that females' judgments show a larger duration judgment ratio (and more inaccuracy) when they produce and verbally estimate durations. This suggests either that, relative to males, females either (1) use units of duration (seconds and minutes) in such a way as to not be in close accord with objective units or (2) attend more to time and thereby experience a lengthened subjective duration. The evidence on the moderating influence of number of trials suggests that the latter explanation is more likely.

Retrospective paradigm. A total of 16 experiments contributed an effect size estimate for the retrospective paradigm. Three variables were significant moderators: number of stimuli, complexity of stimuli, and duration judgment immediacy. For all three moderators, however, one or more classes showed significant heterogeneity of variance; thus, no single variable provided a complete model. There was no significant moderating influence of publication year, modality of stimuli, segmentation of stimuli, duration length, processing type, processing difficulty, or duration judgment method.

A moderator variable may have been significant only because its levels were typically correlated with the levels of another significant moderator. In fact, complexity of stimuli was correlated with duration judgment immediacy [$r(15) = .48, p = .03$]. A multiple regression model containing the three significant moderator variables [$R = .58, Q_R(3) = 21.4, p < .001$] revealed that all were significant predictors: number of stimuli ($\beta = 0.62, p < .001$),

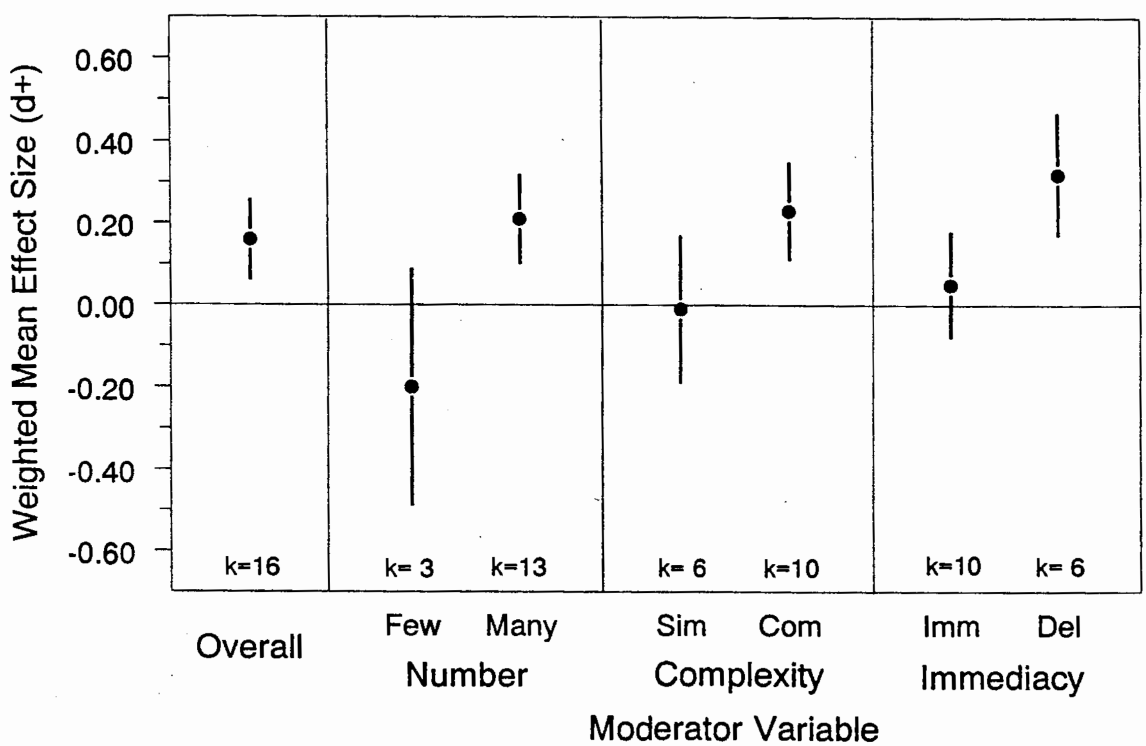

Figure 2. Weighted mean effect size (d_+) and 95% CI in the retrospective paradigm (Overall) and for each class of a moderator variable that a multiple regression analysis revealed to be an important predictor. The number of experimental conditions (*k*) that contributed to each weighted mean is shown. Number = number of stimuli; Few = no stimulus or several stimuli; Many = many stimuli; Complexity = complexity of stimuli; Sim = simple; Com = moderate or complex; Immediacy = duration judgment immediacy; Imm = immediate; Del = delayed.

judgment immediacy ($\beta = 0.45, p < .001$), and complexity of stimuli ($\beta = -0.18, p = .02$). Figure 2 shows results of categorical model testing of the three significant predictors.

Number of stimuli. Number of stimuli showed a between-classes effect [$Q_B(1) = 7.28, p < .01$]: The conditions that contained no stimuli or several stimuli showed no significant effect, whereas those that contained many stimuli showed a positive sex effect.

Complexity of stimuli. Complexity of stimuli showed a between-classes effect [$Q_B(1) = 5.48, p < .05$]. The conditions that contained simple stimuli showed no significant effect, whereas those that contained moderate or complex stimuli showed a positive sex effect.

Duration judgment immediacy. Duration judgment immediacy showed a between-classes effect [$Q_B(1) = 6.52, p < .05$]. The conditions that entailed immediate judgments showed no sex effect, whereas those that entailed delayed judgments showed a positive sex effect.

Duration Judgment Variability

Effect size and primary statistics. Only one experimenter (Seashore, 1899) reported intraindividual variability data. However, a total of 28 experiments, which yielded data from 2,555 females and 2,593 males, contained sufficient information to calculate a coefficient of variation for each sex.[6] We defined the sign of each effect as positive if the coefficient of variation was larger for females and as negative if it was larger for males. The resulting weighted mean effect size ($d_+ = 0.09$, 95% CI = 0.04 to 0.15) indicated a larger coefficient of variation for females than males ($p = .0007$). The unweighted mean effect size was 0.12. The homogeneity statistic revealed that effect sizes were heterogeneous [$Q(27) = 56.7, p = .0003$].

The mean coefficient of variation was significantly larger for females (0.34) than for males (0.30) [$t(27) = 2.21, p = .04$], and the sex ratio (1.15) was significantly greater than 1.00 [$t(27) = 2.35, p = .03$]. Although the effect size was small, the sex ratio showed a moderately large difference: The mean coefficient of variation was 15% larger for females than for males. Considering that the coefficient of variation adjusts for differences in magnitude (i.e., mean duration judgment), the standard deviation was about 29% larger for females than for males.

Moderator analyses. The duration judgment paradigm did not moderate coefficient of variation effect sizes [$Q_B(1) = 0.18, p = 0.67$], so we analyzed other moderator variables without separating them by paradigm. Five moderators were significant: participants' age, processing difficulty, duration length, duration judgment method, and judgment immediacy. The weighted mean effect size was larger in conditions involving children or adolescents, more difficult processing, moderate durations, the production method, and delayed judgments. A multiple regression model containing the five variables accounted for some variability in the effect sizes [$R = .50, Q_R(5) = 8.13, p = .15$]. The only significant moderator was judgment immediacy ($\beta = 0.27, p = .03$). A model containing only judgment immediacy ($\beta = 0.21, p = .04$) successfully accounted for the variability in the effect sizes [$R = .27, Q_R(1) = 4.02, p = .04$].

Psychophysical Slope

Four articles (Carlson & Feinberg, 1970; A. D. Eisler, 1995; A. D. Eisler & H. Eisler, 1994; H. Eisler & A. D. Eisler, 1992) contained data on sex differences in the slope of the psychophysical function relating subjective to objective duration. The three most recent experiments, which used the reproduction method, found no significant sex difference in the slope. The present data are too scarce to draw any conclusion.

SUMMARY AND THEORETICAL DISCUSSION

The meta-analyses reveal several findings on sex differences in duration judgments:

1. The overall meta-analysis reveals a small effect: The ratio of subjective-to-objective duration is about 10% greater for females than for males. Females tend to estimate a subjective duration that is longer than the objective duration, whereas males do not do so significantly. To give an example using conventional units (which reflects the actual mean duration judgment ratios), females verbally estimate a 100-sec duration as 110 sec, but males estimate the duration as 98 sec.

2. The most theoretically important moderator of the overall sex effect is the duration judgment paradigm (prospective vs. retrospective)—that is, whether study participants know in advance that they are required to make a duration judgment.

3. Prospective duration judgments show no significant sex effect. However, prospective judgments are moderated by several variables, especially the number of trials, the participants' age, and the duration judgment method.

4. Retrospective duration judgments show a significant sex effect: The ratio of subjective-to-objective duration is larger for females than for males. Females tend to estimate a subjective duration that is longer than the objective duration, whereas males do not do so significantly. Retrospective judgments are moderated by several variables, especially the number of stimuli presented during the target duration, the complexity of the stimuli, and the delay between the target duration and the judgment.

5. Regardless of the duration judgment paradigm, there is a small sex difference in the intersubject variability of duration judgments: The mean coefficient of variation is 15% larger for females than that for males.

Given these major findings, what possible theoretical interpretations fit the pattern observed? First, the small sex effect in the overall duration judgment magnitude is not surprising. Many sex differences in cognitive processes are small (Halpern, 2000). However, the small overall effect is obscured (or, more properly, is moderated) by the duration judgment paradigm. As such, we do not think that the small size of the overall effect is a concern.

Prospective duration judgments are moderated by three variables: the number of trials, the duration judgment method, and the participants' age. Females may attend to time more frequently than do males during durations that they experience in an experimental context. Attentional models of prospective timing can explain this finding. As the number of duration judgments made in any particular experiment increases, so does the total duration of the experiment. If the total duration of the experiment becomes actually or psychologically longer, a person may attend to time more or less frequently. Our finding suggests that males attend to time less often than do females as the number of trials increases. Duration judgment method is another important moderator. Figure 1 shows that the verbal estimation and production (numerical estimation) methods yield a positive effect size—that is, a larger duration judgment ratio for females than for males. Relative to males, females give larger verbal estimates and make shorter productions. These findings suggest that females focus attention on time more than males do, with the result that they accumulate subjective temporal units at a faster rate. It is unclear whether females make reproductions that are longer or shorter than those of males: Although the effect size statistics show a negative effect (in which females make shorter reproductions), the duration judgment ratios show no significant difference. Finally, the sex difference becomes more positive as the participants' age increases (see Figure 1). The negative effect size for children may be attributable to a correlation with duration judgment method: Most experiments in which children were participants used the reproduction method. No such explanation is possible for the more positive effect for older adults relative to younger adults, which remains unexplained.

The sex effect is positive in the retrospective paradigm: Relative to males, females show a larger ratio of subjective-to-objective duration. Three variables moderate this effect: It is more positive with increases in the number of stimuli, the complexity of stimuli, and the delay between the target duration and the duration judgment. Taken together, these findings suggest that the sex difference in retrospective duration judgments is attributable to memory for events or contextual changes from the target duration. Evidence that females perform relatively better on episodic memory tasks than do males (Herlitz et al., 1999; Herlitz et al., 1997) suggests a simple explanation: Females remember events from the target duration better than do males, and, therefore, they judge the duration as being longer. If more stimuli or more complex stimuli are presented, females may be better able to remember them than may males. Contextual changes are available in memory in association with remembered events. Retrospective duration judgments lengthen as the number of remembered changes in cognitive context increase (Block, 1990). No experiment testing for possible sex differences in manipulated contextual changes per se, and such studies are needed. Females may be relatively more sensitive to contextual changes, or they may encode them in a more permanent way. In addition, the delay between the target duration and the duration judgment is an important moderator. If males forget presented information at a faster rate, they would give a smaller ratio of subjective-to-objective duration than females give. In short, an episodic memory account of remembered duration can explain the sex difference in retrospective duration judgments.

The findings in the retrospective paradigm may also be explained, albeit speculatively, in terms of differences in sex hormones, such as estrogen. Estrogen increases the activity of NMDA, a neurotransmitter found in the hippocampus (Foy et al., 1999). Because hippocampal functioning is critical for the permanent storage of episodic memories, this evidence may help to explain both Herlitz et al.'s (1997) finding that females perform better on episodic memory tasks and our finding that they remember durations as being relatively longer.

Our finding of greater female interindividual variability is the opposite of what we expected. An explanation for the typical finding, as well as for our finding of greater within-sex variability for females than for males, remains elusive. It is somewhat unclear why duration judgment immediacy moderates the sex effect, such that the interindividual variability of females becomes relatively larger than that of males for delayed, relative to immediate, judgments. The finding suggests that, relative to males, different females may use relatively different processes, especially as the delay increases. Because no experiment manipulated judgment immediacy, studies are needed to investigate this finding. Regardless of the possible explanations, our main finding is an important one for theories on within-sex variability in cognitive task performance.

The present evidence either does not support or actually contradicts several possible physiological explanations. These include proposed differences in pacemaker rate, body temperature, or basal metabolism (Hancock, 1993). The finding of a sex difference in the retrospective paradigm, along with the various moderators of duration judgments in both paradigms, is difficult to reconcile with such physiological explanations.

CONCLUSIONS

Our meta-analyses reveal a small but statistically significant sex difference in the magnitude of duration judgments, which is moderated by several factors. There is also a sex difference in the interindividual variability of duration judgments. Attention and memory selectively influence duration judgments in the prospective and retrospective duration judgment paradigms, respectively (Block & Zakay, 1997). Our findings suggest that, relative to males, females sustain attention to time more in the prospective paradigm and have better episodic memory in the retrospective paradigm.

REFERENCES

(References marked with an asterisk [*] indicate studies included in the meta-analyses.)

*Adkins, C. J. (1972). Verbal estimations of time at four spatial distances. *Perceptual & Motor Skills*, **35**, 411-418.

*Allen, D. A. (1980). Filling time versus affective response to the activity that fills the time: Independent effects on time judgment? *Perceptual & Motor Skills*, **51**, 723-727.

*Antick, J. R., & Schandler, S. L. (1993). An exploration of the interaction between variation in wavelength and time perception. *Perceptual & Motor Skills*, **76**, 987-994.

*Axel, R. (1924). Estimation of time. *Archives of Psychology*, **74**, 1-77.

*Baldwin, R. O., Thor, D. H., & Wright, D. E. (1966). Sex differences in the sense of time: Failure to replicate a 1904 study. *Perceptual & Motor Skills*, **22**, 398.

*Barocas, H. A. (1971). Temporal orientation, human movement responses and time estimation. *Journal of Personality Assessment*, **35**, 315-319.

Becker, G. (1996). Bias in the assessment of gender differences [Comment]. *American Psychologist*, **51**, 154-155.

*Bell, C. R., & Watts, A. N. (1966). Personality and judgements of temporal intervals. *British Journal of Psychology*, **57**, 155-159.

Bindra, D., & Waksberg, H. (1956). Methods and terminology in studies of time estimation. *Psychological Bulletin*, **53**, 155-159.

Block, R. A. (1978). Remembered duration: Effects of event and sequence complexity. *Memory & Cognition*, **6**, 320-326.

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., & Eisler, H. (2000). *The compleat bibliography on the psychology of time, 1839–1999* [Machine-readable data file]. Bozeman, MT: Department of Psychology, Montana State University [Producer and distributor].

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, 4, 184-197.

Block, R. A., & Zakay, D. (in press). Psychological time at the millennium: Some past, present, future, and interdisciplinary issues. In M. P. Soulsby & J. T. Fraser (Eds.), *Time: Perspectives at the millennium* (The study of time X). Westport, CT: Greenwood.

Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology & Aging*, **13**, 584-596.

Block, R. A., Zakay, D., & Hancock, P. A. (1999). Developmental changes in human duration judgments: A meta-analytic review. *Developmental Review*, **19**, 183-211.

*Bobko, D. J., Bobko, P., & Davis, M. A. (1986). Effect of visual display scale on duration estimates. *Human Factors*, **28**, 153-158.

Carlson, V. R., & Feinberg, I. (1970). Time judgment as a function of method, practice, and sex. *Journal of Experimental Psychology*, **85**, 171-180.

*Collett, B. A. (1974). Variation in body temperature, perceived duration and perceived personal space. *International Journal of Nursing Studies*, **11**, 47-60.

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

*Cupchik, G. C., & Gebotys, R. J. (1988). The experience of time, pleasure, and interest during aesthetic episodes. *Empirical Studies of the Arts*, **6**, 1-12.

*Davidson, W. B., & House, W. J. (1978). Influence of reflection-impulsivity and cognitive style on time estimation under different ambient conditions. *Perceptual & Motor Skills*, **46**, 1083-1091.

*DeLay, E. R., & Richardson, M. A. (1981). Time estimation in humans: Effects of ambient illumination and sex. *Perceptual & Motor Skills*, **53**, 747-750.

*Dinges, N. G., Tollefson, G., Parks, G. A., & Hollenbeck, A. R. (1978). Anticipated reward and time estimation in young Navajo children. *Perceptual & Motor Skills*, **47**, 1011-1014.

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

*Dubey, R. S. (1978). Effect of personal factors on the estimation of a short-filled time-interval of ten minutes. *Indian Psychological Review*, **16**, 29-35.

*Dubey, R. S. (1986). Effect of complexity, familiarity, anxiety, and sex on the estimation of a series of five short-filled time-intervals. *Indian Journal of Behaviour*, **10**, 22-31.

*Dubey, R. S., & Sharma, S. C. (1978). The effect of manifest anxiety on time-estimation. *Journal of Psychological Researches*, **22**, 153-164.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, **110**, 109-128.

*Eisler, A. D. (1995). Cross-cultural differences in time perception: Comparison of African immigrants and native Swedes. In G. Neely (Ed.), *Perception and psychophysics in theory and application* (pp. 137-145). Stockholm: Stockholm University.

*Eisler, A. D., & Eisler, H. (1994). Subjective time scaling: Influence of age, gender, and type A and type B behavior. *Chronobiologia*, **21**, 185-200.

*Eisler, H., & Eisler, A. D. (1992). Time perception: Effects of sex and sound intensity on scales of subjective duration. *Scandinavian Journal of Psychology*, **33**, 339-358.

*Fedotchev, A. I. (1984). [The perception, reproduction, and estimation of time in school children]. *Voprosy Psikhologii*, **1**, 80-86.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, **62**, 61-84.

*Filer, R. J., & Meals, D. W. (1949). The effect of motivating conditions on the estimation of time. *Journal of Experimental Psychology*, **39**, 327-331.

Foy, M. R., Xu, J., Xie, X., Brinton, R. D., Thompson, R. F., & Berger, T. W. (1999). 17beta-estradiol enhances NMDA receptor-mediated EPSPs and long-term potentiation. *Journal of Neurophysiology*, **81**, 925-929.

Fraisse, P. (1998). Time perception. In *The new encyclopædia Britannica* (15th ed., Vol. 25, pp. 494-496). Chicago: Encyclopædia Britannica.

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

*Gilbert, J. A. (1894). Researches on the mental and physical development of school-children. *Studies From the Yale Psychology Laboratory*, **3**, 40-100.

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

Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.

*Gulliksen, H. (1927). The influence of occupation upon the perception of time. *Journal of Experimental Psychology*, **10**, 52-59.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist*, **52**, 1091-1102.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Erlbaum.

Hancock, P. A. (1993). Body temperature influence on time perception. *Journal of General Psychology*, **120**, 197-216.

Hancock, P. A. (1999). *Gaia and Chronos: Sex differences in the fourth cognitive dimension* (Tech. Rep. No. 99-01). Minneapolis, MN: University of Minnesota, Human Factors Research Laboratory.

*Hancock, P. A., Arthur, E. J., Chrysler, S. T., & Lee, J. (1994). The effects of sex, target duration, and illumination on the production of time intervals. *Acta Psychologica*, **86**, 57-67.

*Hancock, P. A., & Manser, M. P. (1997). Time to contact: More than tau alone. *Ecological Psychology*, **9**, 265-297.

*Hancock, P. A., Vercruyssen, M., & Rodenburg, G. J. (1992). The effect of gender and time-of-day on time perception and mental workload. *Current Psychology: Research & Reviews*, **11**, 203-225.

*Hawkins, N. E., & Meyer, M. E. (1965). Time perception of short intervals during finished, unfinished and empty task situations. *Psychonomic Science*, **3**, 473.

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., & NOWELL, A. (1995). Sex differences in mental scores, variability, and numbers of high-scoring individuals. *Science*, **269**, 41-45.

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

HERLITZ, A., AIRAKSINEN, E., & NORDSTROEM, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology*, **13**, 590-597.

HERLITZ, A., NILSSON, L.-G., & BÄCKMAN, L. (1997). Gender differences in episodic memory. *Memory & Cognition*, **25**, 801-811.

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

*HICKS, R. E., ALLEN, D. A., & MAYO, J. P. (1984). A developmental study of temporal duration judgments. *Journal of Genetic Psychology*, 144, 31-38.

*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.

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

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

JOHNSON, B. T., & TURCO, R. M. (1992). The value of goodness-of-fit indices in meta-analysis: A comment on Hall and Rosenthal. *Communication Monographs*, **59**, 388-396.

*KELLARIS, J. J., & ALTSECH, M. B. (1992). The experience of time as a function of musical loudness and gender of listener. *Advances in Consumer Research*, **19**, 725-729.

*KELLARIS, J. J., & MANTEL, S. P. (1994). The influence of mood and gender on consumers' time perceptions. *Advances in Consumer Research*, **21**, 514-518.

*KIRKCALDY, B. D. (1984). Individual differences in time estimation. *International Journal of Sport Psychology*, **15**, 11-24.

*KLINE, D. W., & BURDICK, D. C. (1980). Metronome pacing and age: The effect of irrelevant stimuli. *Experimental Aging Research*, **6**, 393-398.

*KLINE, D. W., HOLLERAN, L., & ORME-ROGERS, C. (1980). The estimation of short time intervals as a function of age and metronome pacing. *International Journal of Aging & Human Development*, **11**, 127-134.

*KRISHNA, K. P., & SINHA, J. K. (1974). Personality adjustment and time perception in normal adolescents. *Indian Journal of Applied Psychology*, **11**, 24-27.

*KURZ, R. B., COHEN, R., & STARZYNSKI, S. (1965). Rorschach correlates of time estimation. *Journal of Consulting Psychology*, **29**, 379-382.

*LANDAETA, F. S., SAAVEDRA, M. A., & SIMICIC, V. Y. (1981). La percepción del tiempo y su relación con la edad [Time perception and its relation to age]. *Revista Chilena de Psicología*, 4, 35-41.

*LEVITT, E. E. (1983). Estimating the duration of sexual behavior: A laboratory analog study. *Archives of Sexual Behavior*, **12**, 329-335.

LINN, M. C., & PETERSEN, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, **56**, 1479-1498.

*LOFTUS, E. F., SCHOOLER, J. W., BOONE, S. M., & KLINE, D. (1987). Time went by so slowly: Overestimation of event duration by males and females. *Applied Cognitive Psychology*, **1**, 3-13.

MACCOBY, E. E., & JACKLIN, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

*MACDOUGALL, R. (1904). Sex differences in the sense of time. *Science*, **19**, 707-708.

*MARTIN, G. A., SHUMATE, M., & FRAUENFELDER, K. (1981). Experience of duration as a function of number of responses, task difficulty, and sex. *Perceptual & Motor Skills*, **53**, 139-145.

*MATSUDA, F. (1965a). [Development of time estimation: I. Effects of degree of unity in verbal auditory stimulus]. *Japanese Journal of Psychology*, **36**, 169-177.

*MATSUDA, F. (1965b). [Development of time estimation: II. Effects of frequency of sounds given during standard time]. *Japanese Journal of Psychology*, **36**, 285-294.

*MATSUDA, F. (1967). [Development of time estimation: III. Effects of frequency of sounds given during standard time and reproduced time]. *Japanese Journal of Psychology*, **37**, 352-358.

*MATSUDA, F., & MATSUDA, M. (1974). Effects of frequency of intermittent stimuli on time estimation in children and in adults: I. Sounds and lights. *Psychologia: An International Journal of Psychology in the Orient*, **17**, 206-212.

*MONTARE, A. (1985). Learning effects of knowledge of results upon time estimation. *Perceptual & Motor Skills*, **60**, 871-877.

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

*MONTGOMERY, J. D. (1979). Variations in perception of short time intervals during menstrual cycle. *Perceptual & Motor Skills*, **49**, 940-942.

MORSE, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology & Aging*, **8**, 156-164.

*MYERS, G. C. (1916). Incidental perception. *Journal of Experimental Psychology*, **1**, 339-350.

*NAIL, P., LEVY, L., RUSSIN, R., & CRANDALL, R. (1981). Time estimation and obesity. *Personality & Social Psychology Bulletin*, 7, 139-146.

*NEWMAN, M. A. (1982). Time as an index of expanding consciousness with age. *Nursing Research*, **31**, 290-293.

*POLLACK, I. W., OCHBERG, F. M., & MEYER, E. (1969). Social class and the subjective sense of time. *Archives of General Psychiatry*, **21**, 1-14.

*POLYUKHOV, A. M. (1989). Subjective time estimation in relation to age, health, and interhemispheric brain asymmetry. *Zeitschrift für Gerontologie*, **22**, 79-84.

POYNTER, W. D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 305-331). Amsterdam: North-Holland.

*RAMMSAYER, T. H. (1998). Temporal information processing in male and female subjects. *Studia Psychologica*, **40**, 149-164.

*ROECKELEIN, J. E. (1972). Sex differences in time estimation. *Perceptual & Motor Skills*, **35**, 859-862.

*SAMUELSON, D. J., & LINDAUER, M. S. (1976). Perception, evaluation, and performance in a neat and messy room by high and low sensation seekers. *Environment & Behavior*, **8**, 291-306.

*SARASON, I. G., & STOOPS, R. (1978). Test anxiety and the passage of time. *Journal of Consulting & Clinical Psychology*, **46**, 102-109.

*SCHIFFMAN, H. R., & BOBKO, D. J. (1974). Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology*, **103**, 156-159.

*SEASHORE, C. E. (1899). Some psychological statistics. *University of Iowa Studies in Psychology*, **2**, 1-84.

SETH-SMITH, M., ASHTON, R., & MCFARLAND, K. (1989). A dual-task study of sex differences in language reception and production. *Cortex*, **25**, 425-431.

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.

*SIMPSON, M. L. (1982). Time-of-day effects in performance by seventh grade students on two measures of impulse control. *Perceptual & Motor Skills*, **55**, 115-121.

*SMITH, M. J. (1975). Changes in judgment of duration with different patterns of auditory information for individuals confined to bed. *Nursing Research*, **24**, 93-98.

*SMITH, M. J. (1979). Duration experience for bed-confined subjects: A replication and refinement. *Nursing Research*, **28**, 139-144.

*SMITH, M. J. (1984). Temporal experience and bed rest: Replication and refinement. *Nursing Research*, **33**, 298-302.

*SMITH, N. C., JR. (1969). The effect on time estimation of increasing the complexity of a cognitive task. *Journal of General Psychology*, **81**, 231-235.

*SRIVASTAVA, R., & SRIVASTAVA, B. (1978). Time estimation as a function of religion, sex and anxiety. *Psychological Studies*, **23**, 122-124.

STRANG, H. R., RUST, J. O., & GARRISON, G. (1973). Sex differences in short-term time estimation. *Perceptual & Motor Skills*, **36**, 1109-1110.

*SWIFT, E. J., & MCGEOCH, J. A. (1925). An experimental study of the perception of filled and empty time. *Journal of Experimental Psychology*, **8**, 240-249.

*TROUTWINE, R., & O'NEAL, E. C. (1981). Volition, performance of a boring task and time estimation. *Perceptual & Motor Skills*, **52**, 865-866.

*WARNER, D. J., & BLOCK, R. A. (1984). Type A behavior and temporal judgment. *Bulletin of the Psychonomic Society*, **22**, 163-166.

WOODROW, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224-1236). New York: Wiley.

*YARMEY, A. D. (1990). Accuracy and confidence of duration estimates following questions containing marked and unmarked modifiers. *Journal of Applied Social Psychology*, **20**, 1139-1149.

*YERKES, R. M., & URBAN, F. M. (1906). Time-estimation in its relations to sex, age, and physiological rhythms. *Harvard Psychological Studies*, **2**, 405-430.

ZAKAY, D. (1990). The evasive art of subjective time measurement: Some methodological dilemmas. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 59-84). Hillsdale, NJ: Erlbaum.

*ZAKAY, D. (1992a). On prospective time estimation, temporal relevance and temporal uncertainty. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 109-117). Dordrecht: Kluwer.

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

*ZAKAY, D. (1993). The roles of non-temporal information processing load and temporal expectations in children's prospective time estimation. *Acta Psychologica*, **84**, 271-280.

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

NOTES

1. We included any relevant article listed in PsycINFO or Medline as of December 1999.

2. Strang, Rust, and Garrison (1973) analyzed data on each participant's coefficient of variation, finding that intraindividual consistency was greater for males than for females (i.e., the coefficient of variation was smaller for males than for females).

3. We converted quartile deviations (*QD*s) to standard deviations (*SD*s) by using the approximation $SD = 1.4826\ QD$ and absolute deviations (*AD*s) to standard deviations by using the approximation $SD = 1.2533\ AD$ (Guilford, 1936).

4. We repeated all moderator variables analyses (which we report in subsequent parts of the Results section), excluding experimental conditions for which insufficient data were reported to calculate an exact d value (i.e., those for which we had to estimate $d = 0$). When we did so, the basic findings did not change much, although they usually intensified slightly: Nonsignificant findings remained nonsignificant, and significant findings became more significant.

5. This is also the main reason why the overall weighted mean effect did not change much when we included only studies for which we were able to calculate an exact effect size, because studies for which we had to estimate $d = 0$ received less weight in the former analysis than did studies for which $d \neq 0$.

6. The Appendix indicates these studies. A table showing data from each included study is available from the first author.

APPENDIX
Duration Judgment Ratios and Effect Sizes in Each Paradigm (Ordered by Publication Year)

Experiment	Female		Male		Female-to-Male		Effect Size (*d*)
	Ratio	*n*	Ratio	*n*	Ratio	*N*	
Prospective Paradigm							
Gilbert (1894, Test 11)†	0.78	536	0.80	548	0.97	1,084	−0.32
Seashore (1899)[a]†	0.92	24	0.98	20	0.94	44	−0.25
MacDougall (1904)[b]	1.82	15	1.20	15	1.52	30	—
Yerkes & Urban (1906)[c]†	0.77	20	0.99	20	0.78	40	−0.86
Yerkes & Urban (1906)[d]†	1.03	274	0.85	251	1.21	525	0.50
Axel (1924)†	1.10	22	0.73	18	1.50	40	0.86
Swift & McGeoch (1925)[e]†	1.53	104	1.55	123	0.98	227	−0.04
Swift & McGeoch (1925)[f]†	1.01	550	0.99	631	1.02	1,181	0.05
Gulliksen (1927)†	1.07	154	0.97	172	1.10	326	0.22
Hawkins & Meyer (1965)	—	9	—	9	—	18	−0.31
Kurz, Cohen, & Starzynski (1965)	—	25	—	26	—	51	0.00*
F. Matsuda (1965a)†	0.75	32	0.81	32	0.93	64	−0.24
F. Matsuda (1965b, Experiment 1)	—	32	—	32	—	64	0.00*
F. Matsuda (1965b, Experiment 2)	—	32	—	32	—	64	0.00*
Baldwin, Thor, & Wright (1966)	—	78	—	40	—	118	−0.12
Bell & Watts (1966, Investigation V)	—	22	—	13	—	35	0.00*
F. Matsuda (1967)	—	32	—	32	—	64	0.00*
Hornstein & Rotter (1969)	1.09	18	0.95	18	1.14	36	0.39
Pollack, Ochberg, & Meyer (1969)	—	21	—	19	—	40	0.26
N. C. Smith (1969)	—	16	—	16	—	32	0.00*
Barocas (1971)	—	11	—	29	—	40	0.00*
Adkins (1972, Experiment 1)	1.01	20	0.79	20	1.29	40	0.92
Roeckelein (1972)	0.85	15	0.84	15	1.01	30	0.00
Collett (1974)†	0.99	50	1.00	50	0.99	100	−0.04
Krishna & Sinha (1974)	1.42	34	1.37	50	1.03	84	0.11
F. Matsuda & M. Matsuda (1974)	—	30	—	30	—	60	0.00*
Schiffman & D. J. Bobko (1974)	—	27	—	36	—	63	0.04
M. J. Smith (1975)	—	90	—	90	—	180	0.00*
Hicks, Miller, & Kinsbourne (1976)[g]	—	30	—	30	—	60	0.00*
Davidson & House (1978)†	1.01	64	0.99	34	1.02	98	0.11
Dinges, Tollefson, Parks, & Hollenbeck (1978)†	0.71	56	0.62	56	1.14	112	0.18
Dubey (1978)	0.91	36	0.99	34	0.93	70	−0.36
Sarason & Stoops (1978, Experiment 1)	—	48	—	48	—	96	0.00*
Sarason & Stoops (1978, Experiment 2)	—	60	—	60	—	120	0.00*
R. Srivastava & B. Srivastava (1978)	0.76	160	0.89	160	0.85	320	−0.63
Montgomery (1979)†	1.28	10	0.92	10	1.40	20	0.67
M. J. Smith (1979)	—	60	—	60	—	120	0.00*
Allen (1980)	—	12*	—	12*	—	24*	0.00*
D. W. Kline & Burdick (1980)	—	15	—	15	—	30	0.00*
D. W. Kline, Holleran, & Orme–Rogers (1980)	0.99	12	0.83	12	1.19	24	0.27*
Delay & Richardson (1981)	0.98*	7	0.97*	7	1.02*	14	0.34*
Landaeta, Saavedra, & Simicic (1981)	—	30	—	30	—	60	0.00*
Troutwine & O'Neal (1981)	—	20	—	20	—	40	0.00*
Newman (1982)	1.51	68	1.21	11	1.26	79	0.67
Simpson (1982)†	1.41	52	1.18	52	1.20	104	0.56
Fedotchev (1984)†	0.94	34	0.90	27	1.04	61	0.04
Hicks, Allen, & Mayo (1984)	—	45	—	45	—	90	0.00*
Kirkcaldy (1984)†	1.15	61	1.04	44	1.10	105	0.30
M. J. Smith (1984)	—	60	—	60	—	120	0.00*
Warner & Block (1984)	—	56	—	56	—	112	0.00*
Zakay & Fallach (1984, Experiment 1)	—	68	—	21	—	89	0.00*
Zakay & Fallach (1984, Experiment 2)	—	22	—	8	—	30	0.00*
Zakay & Fallach (1984, Experiment 3)[g]	—	20	—	20	—	40	0.00*
Montare (1985, Experiment 1)†	1.09	12	0.94	12	1.16	24	0.49
Dubey (1986)	1.05	9	0.95	9	1.10	18	0.77
Cupchik & Gebotys (1988)	—	16	—	16	—	32	0.00*
Montare (1988, Experiment 1)[h]	—	20	—	20	—	40	0.00*
Polyukhov (1989)†	0.95*	162	0.98*	94	0.97*	256	−0.09*
H. Eisler & A. D. Eisler (1992)	0.95*	6	0.84*	6	1.13*	12	1.37
Hancock et al. (1992, Experiment 1)	1.08	12	0.91	12	1.19	24	2.52*

APPENDIX (Continued)

Experiment	Female Ratio	Female n	Male Ratio	Male n	Female-to-Male Ratio	Female-to-Male N	Effect Size (d)
Hancock et al. (1992, Experiment 2)	0.97*	12	0.81*	12	1.19*	24	2.52*
Zakay (1992a, Experiment 1)	—	63	—	63	—	126	0.00*
Zakay (1992a, Experiment 2)	—	80	—	80	—	160	0.00*
Zakay (1992b, Experiment 1)[g]	—	28	—	28	—	56	0.00*
Zakay (1992b, Experiment 2)	—	40	—	40	—	80	0.00*
Antick & Schandler (1993, Study One)	—	23*	—	22*	—	45	0.00*
Antick & Schandler (1993, Study Two)	—	9*	—	9*	—	18	0.00*
Zakay (1993)	—	36	—	36	—	72	0.00*
A. D. Eisler & H. Eisler (1994)†	0.84	24	0.77	24	1.09	48	0.60
Hancock et al. (1994)†	1.09	6	0.84	6	1.31	12	1.30
A. D. Eisler (1995)	0.78*	14	0.70*	14	1.11*	28	1.21
Hancock & Manser (1997, Experiment 1)	—	20	—	20	—	40	0.00*
Hancock & Manser (1997, Experiment 2)	—	12	—	12	—	24	0.00*
Rammsayer (1998, Experiment 1)†	0.56*	17	0.76*	17	0.74*	34	−1.28
Rammsayer (1998, Experiment 3)†	0.65*	16	0.81*	16	0.81*	32	−1.13
Overall mean (unweighted)[i]	1.04*	74	0.96*	74	1.09*	74	0.16*
Overall mean (unweighted)[j]	1.05*	37	0.96*	37	1.09*	37	0.32*
Overall mean (weighted)[k]		4,021*		3,872*		7,893*	0.02*
Overall mean (weighted)[l]		2,823*		2,736*		5,549*	0.03*
Retrospective Paradigm							
Myers (1916, Experiment 3)†	2.38	32	1.62	68	1.73	100	0.93
Filer & Meals (1949)[m]	1.06	31	1.03	29	1.03	60	0.07
Hicks et al. (1976)[g]	—	30	—	30	—	60	0.00*
Samuelson & Lindauer (1976, Study 2)	0.89	28	1.18	28	0.76	56	−0.62
Dubey & Sharma (1978)†	1.08	90	1.03	90	1.05	180	0.58
Martin, Shumate, & Frauenfelder (1981)†	0.83	54	0.77	40	1.07	94	0.21
Nail, Levy, Russin, & Crandall (1981)†	0.87	52	1.01	46	0.85	98	−0.41
Levitt (1983)[n]†	1.46	11	1.12	56	1.30	67	0.74
Zakay & Fallach (1984, Experiment 3)[g]	—	20	—	20	—	40	0.00*
D. J. Bobko, P. Bobko, & Davis (1986)†	1.84	36	1.36	36	1.35	72	0.45
Loftus, Schooler, Boone, & D. Kline (1987, Experiment 2)	5.50	156	4.44	110	1.24	266	0.28
Loftus et al. (1987, Experiment 3)	3.01	69	2.04	68	1.47	137	0.53
Yarmey (1990)	—	45	—	40	—	85	0.00*
Kellaris & Altsech (1992)	0.92	54	1.09	54	0.85	108	−0.46
Zakay (1992b, Experiment 1)[g]	—	28	—	28	—	56	0.00*
Kellaris & Mantel (1994)	0.75	37*	0.89	73*	0.85	110	−0.42
Overall mean (unweighted)[i]	1.72	16	1.46	16	1.11	16	0.12*
Overall mean (unweighted)[j]	1.72	12	1.46	12	1.11	12	0.16
Overall mean (weighted)[k]		773		816		1,589	0.16*
Overall mean (weighted)[l]		650		698		1,348	0.19*
Overall (Both Paradigms)							
Overall mean (unweighted)[i]	1.21*	87	1.06*	87	1.10*	87	0.16*
Overall mean (unweighted)[j]	1.20*	51	1.09*	51	1.09*	51	0.27*
Overall mean (weighted)[k]		4,794*		4,688*		9,482*	0.03*
Overall mean (weighted)[l]		3,473*		3,424*		6,880*	0.06*

Note—Positive effect size (d) indicates that the duration judgment ratio was larger for females than for males; negative effect size (d) indicates that it was larger for males than for females. A dash indicates that the article did not contain sufficient data. *An approximate datum (e.g., one estimated from a figure). †Experimental report also provided coefficient of variation data. [a]We list this experiment only for the sake of completeness; these data were not included in the effect size analyses. [b]We calculated from data in Table XVII. [c]We calculated from the data on page 418. [d]We calculated from the data on page 426. [e]We calculated from the data in Table I. [f]We calculated from the data in Table VII. [g]Experiment used both prospective and retrospective paradigms. [h]We only used the data from Experiment 1 because Experiments 2 and 3 involved repeated measures. [i]Each mean weights each experiment equally, and for d this includes those not reporting duration judgment means. [j]Each mean weights each experiment equally, but only experiments for which we could calculate an exact (i.e., nonzero) d. [k]Each mean weights each experiment by n contributing to each datum (and by TW for effect size). [l]Each mean weights each experiment by n contributing to each datum (and by TW for effect size), but only experiments for which we could calculate an exact (i.e., nonzero) d. [m]We calculated from the control group data. [n]We calculated from the data of the 1978 sample.

(Manuscript received September 13, 1999;
revision accepted for publication February 9, 2000.)