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 ZARAY  
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 4,988 males yielded 87 effect size estimates of magnitude and 26 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 interobserver variability than did males’ judgments. Relative to males, females sustain attention to some 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, faces, and performed activities, than do males (Herlitz, Asakishin, & Nordström, 1999; Herlitz, Nils-son, & Bäckman, 1977). 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., 1977).

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 adjoins and represents temporal properties in order to synchronize action with external events. For example, while driving down 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 MONTES 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 1999. The authors thank John Hall for reviewing some articles, (in) Patricia, Dida Sensor, and (in) Fukao Manri for reviewing some articles, and (in) Arthur for coding some moderators. The authors also thank Scott Brown, Alice Eagle, Hannah Evidel, Diane Halpern, Charles Pears, 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@mont- state.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 psycho-physical investigations (Woodworth, 1938). 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 (McDougall, 1904; Seashott, 1899; Yerkes & Utrian, 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 & McNeroy, 1952). Researchers offered 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. Eiker & A. D. Eiker, 1992; Hancock, Arthur, Chyssler, & Lee, 1994; Hancock, Verrryjohnsen, & Rodenburg, 1992). Others thought that the literature showed that females underestimate duration more than do males (Kellaris & Mustel, 1994). Explanations for these more recent findings remain elusive. One eminent reviewer concludes: "Sex differences [in estimates of duration] have not been reliably established" (Fisler, 1998, p. 496). Another researcher (Hancock, 1990), 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 spatial-temporal 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 construct 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

Unsurprisingly, the most important variable influencing duration judgments is the actual target duration. Arguably, the next most important variable influencing duration judgments is whether a person knows in advance not a duration estimate will be required. In the prospective paradigm, a person has the 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 judgments are 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 no 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., Her- litz et al., 1975; Seif-Smith, Ashton, & McFaulan, 1980). If this is the case, we expect to find little sex difference in the prospective paradigm.

In the retrospective paradigm, duration judgments are directly related to the amount of encoded and retrieved memory information, such as encoding events or contextual changes. Remembered duration decreases if events change are fewer or more difficult to remember (Block, 1973, 1985). 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., 1990, 1997). Thus, we expect to find a sex difference in retrospective duration judgments, with females remembering durations at 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 real-time 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 you tell me to stop"). Here, we combine data from 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 timekeeping processes, such as stimulant drugs, e.g., Frankenhuysen, 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 experience of 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 effects across experimental conditions. The outcomes clearly concerns 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 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 & Osaka, 2001). It includes articles from Psychological Science, Cognition, and from Medicine and Medici (1966-1999), using the keywords "time perception." It also includes other references (see Block & Zakay, 1997; for details). We searched Social Sciences Citation Index (Social Science, 1977-1999) for articles that cited relevant articles.

To be included in our meta-analyses, experiments 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, human beings without gross psychopathology, or participants experiencing an unusual physical condition or an altered state of consciousness. We also excluded (1) studies that contained only absolute-error data, percentage underestimation and overestimation, and other such measures, (2) experiments involving other temporal judgments, such as speed or discrimination, (3) experiments using nonconventional 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 sex as a factor in the results analyses. We did not include such experiments in our meta-analyses.

Effect Size Analyses

The 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). Wherever 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 judgments for males or females, or if it was not ascertainable. If we could not determine the direction of the effect, we assumed $p = .50$ and calculated the effect size accordingly.

Coded Variables and Moderator Analyses

We tested the homogeneity of each set of $t$-tests to determine whether the conditions allowed a common effect size. We attempted to extend heterogeneity of effect sizes on indicated by the $Q$ statistic.
in terms of coded or manipulated attributes. Continuous moderator variables were tested by using either a weight-by-levels 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 sim- ilar classes of a variable if there were fewer than three effect size esti- mates in a given class. These techniques yield a between-classes effect size, revealing whether the variable is a significant moderator of the effect size. Finally, we assessed the relative importance of each significant moderator variable by relying on several statistics: (1) the results of an analysis of variance 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 indi- cated by each within-class heterogeneity of variance (Johnson & Turco, 1992).

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

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

1. Publication year. This was the copyright date of the article.
2. Participant age. This was coded based on the age range (or other such information) reported: children (8.0-12.9 years old); adolescents (12.0-17.9 years old); young adults, as well as adults, such as all studies involving college students (18.0-29.9 years old); older adults (50.0- 69.9 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 in- formed in advance that they would make judgments about how long, if they made more than one judgment) or as retrospective (if the de- scriptor implied that participants are in- formed in advance that they made only one judgment).
4. Stimulus. This was coded as non (i.e., an "empty" duration, one not containing any presented stimulus), one (e.g., a continuous tone), several (e.g., three words), or voice (e.g., a film or a passage of text).
5. Modality of stimulus. This was the modality of any presented information: visual, verbal, or auditory (e.g., words on a screen, auditory, e.g., a series of tones, voice, or music; e.g., a film with sound and video).
6. Duration length. This was coded as short (i.e., the basis of the predominant length of the target duration, as very short (less than 5.0 sec), short (5.0-14.9 sec), medium (15.0-59.9 sec), or long (60.0 sec or longer).
7. Processing type. This was coded as rapid and accurate (e.g., listening to a text), accurate and veridical (e.g., listening to a series of tones, voice, or music), or fast and accurate (e.g., listening to a series of tones, voice, or music).
8. Processing difficulty. This was coded as easy (e.g., listening to a tone, voice, or music), difficult (e.g., performing a Sperling test), or difficult (e.g., making semantic decisions about words categorized).
9. Duration judgment method. This was coded as verbal estimation, production, reproduction, or reproduction of other text category.
10. Duration judgment format. This was coded as immediate or 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).
11. Number of trials. This was the total number of duration judg- ments made by each participant in the experiment (or in an experi- mental condition). Two of us coded the experiment attributes independently, resolving disagreements by discussion. For some variables, we used the categorical model, unless, or not applicable (e.g., if no stimulus was presented, complexity of stimuli is not applicable). The moder- ator analyses included 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 rep- resented; 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 re- searchers report this measure (see Blesch, Zawacki, & Hanceko, 1964, 1999). For the verbal estimation method, this is the ratio of the per- son's numerical (subjective) response to the target (objective) dura- tion. For the production method, this 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 judgments ratio, hereafter called the sex ratio. These data were available from 60 of 87 (71%) of the experiments. We cautiously used these statistics to examine gender differences in the determinants of duration judgment dura- tion ratios across moderator variable levels. We first report results for estimates. We then report results for the sex ratio, and in a separate analysis, we report results for the sex ratio. Finally, we reported results for the sex ratio. We also reported results for the sex ratio.
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 < 0.01). The duration judgment ratio for females (1.21) was significantly greater than 1.00 (t(48) = 2.01, p = 0.05), whereas that for males (1.06) was not significantly different from 1.00 (t(48) = 1.96, p = 0.06). The sex ratio (1.10) was significantly greater than 1.00 (t(47) = 3.32, p < 0.001). 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 these variables 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 (Qb(1) = 6.77, p = 0.01). 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 fine moderators, however, one or more classes showed significant heterogeneity of variance; thus, no single variable provided a complete model (Johnson & Turco, 1992). Several other co-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 = 0.37, p < 0.03) and duration judgment method (r = 0.25, p < 0.01). Participants' age was correlated with duration judgments.
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 \( R^2 = .56, Q(3) = 36.8, p < .001 \): number of trials (standardized regression coefficient \( \beta = 0.11, p = .011 \)), 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 \( Q(2) = 39.2, p < .0001 \), 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 intradividual 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 \( r(34) = 2.08, p < .05 \). 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](image-url)
classes. Participants' age showed a significant between-class effect ($Q_M(2) = 21.8, p < .001$). 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-class effect ($Q_M(1) = 21.8, p < .001$). 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 ratio suggests that females (0.93) and males (0.89) underestimate durations to the same extent. Females significantly underperformed durations (1.13), but males do not (1.00). Females verbally overestimate durations (1.16), 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 ($r(15) = .48, p = .33$). A multiple regression model containing the three significant moderator variables ($F = 5.9, Q_M(3) = 21.4, p < .001$) revealed that all were significant predictors: number of stimuli ($\beta = .62, p < .001$),

![Figure 2. Weighted mean effect size (d) and 95% CI in the retrospective paradigm (Overall) and for each class of a moderator variable (not 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 stimuli or several stimuli; Many = many stimuli; Complexity = complexity of stimuli; Sim = simple; Com = moderate or complex; Imm = immediate; Del = delayed.)](image-url)
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 ($Q_d(1) = 7.28, p < .01$). 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. Complex stimuli showed a within-classes effect ($Q_d(1) = 5.68, p < .05$). 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 ($Q_d(1) = 6.52, p < .05$). 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 intra-individual variability data. However, a total of 28 experiments, which yielded data from 2.555 females and 2.592 males, contained sufficient information to calculate a coefficient of variation for each sex. We defined the sign of each effect as positive if the coefficient of variation was larger for males and as negative if it was larger for females. 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 ($p = 0.0007$). The unweighted mean effect size was 0.12. The homogeneity statistic revealed that effect sizes were heterogeneous ($Q_{d-27} = 56.7, p = 0.0003$).

The mean coefficient of variation was significantly larger for females (0.34) than for males (0.30) ($t_{27} = 2.21, p = 0.04$), and the sex ratio (1.15) was significantly greater than 1.00 ($t_{27} = 2.35, p = 0.03$). 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. Testing that the coefficients of variation are equal for females and males, the standard deviation of the sex ratio was 0.29% larger for males than for females. Moderators analyses. The duration judgment paradigm did not moderate coefficient of variation effect sizes ($Q_d(1) = 0.18, p = 0.67$), so we analyzed other moderators 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 judgment. A multiple regression model containing the five variables accounted for some variability in the effect sizes ($R^2 = 0.50, Q_{d-5} = 6.13, p = 0.15$). The only significant moderator was judgment immediacy ($\beta = 0.27, p = 0.03$). A model containing only judgment immediacy ($\beta = 0.21, p = 0.04$) successfully accounted for the variability in the effect sizes ($R^2 = 0.27, Q_d(1) = 4.02, p = 0.04$).

**Psychophysical Slope**

Four articles (Carlson & Feinberg, 1970; A. D. Eyster, 1995; A. D. Eyster & H. Eyster, 1994; Eyster & A. D. Eyster, 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 same method, found no significant sex difference in the slope. The present data are too scarce to draw any conclusions.

**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 90% 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 ratio), 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, retrospective 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 inter-subject variability of duration judgments. The mean coefficient of variation is 15% larger for females than 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 pro-
cesses are small (Hampton, 2000). However, the small
overall effect is observed (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 new 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 par-
ticular 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 fe-
males as the number of trials increases. Duration judg-
ment method is another important moderator. Figure 1
shows that the verbal estimation and production (numer-
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 fe-
males 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 repro-
ductions), the duration judgment ratios show no signifi-
cant 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
production 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 ef-
fct: 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 to-
gether, these findings suggest that the sex difference in
retrospective duration judgments is attributable to mem-
ory for events or contextual changes from the target du-
ratio. 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: Fe-
males 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 males. Contextual changes are available in mem-
ory in association with remembered events. Retrospec-
tive duration judgments lengthen as the number of remem-
bered changes in cognitive context increase (Block, 1990).
No experiment testing for possible sex differences in ma-
ipulated contextual changes per se, and such studies are
needed. Females may be relatively more sensitive to con-
textual changes, or they may encode them in a more per-
manent way. In addition, the delay between the target
duration and the duration judgment is an important mod-
erator. 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 VMH, a neurotransmitter found in the hip-
pcocampi (Foy et al., 1999). Because hippocampal func-
tioning is critical for the permanent storage of episodic
memories, this evidence may help to explain both Her-
litz et al.’s (1997) finding that females perform better on
episodic memory tasks and our finding that they remem-
ber durations as being relatively longer.

Our finding of greater female interindividual variabil-
ity 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 inter-
individual variability of females becomes relatively larger
than that of males for delayed, relative to immediate, judg-
ments. The finding suggests that, relative to males, dif-
ferent females may use relatively different processes, es-
pecially as the delay increases. Because no experiment
manipulated judgment immediacy, studies are needed to
investigate this finding. Regardless of the possible ex-
planations, our main finding is an important one for the-
ozes on within-sex variability in cognitive task perfor-

The present evidence either does not support or actu-
ally contradicts several possible physiological explanati-
ions. These include proposed differences in pacemaker
rate, body temperature, or basal metabolism (Slanock,
1993). The finding of a sex difference in the retrospective
paradigm, among 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 sig-
nificant 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 ret-
rospective duration judgment paradigms, respectively
(Block & Eales, 1997). Our findings suggest that, rela-
tive to males, females sustain attention to time more in
SEx AND DURATION JUDMENTS


**Notes**

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

2. Steng, Ross, and Gaskin (1972) analyzed data on each partici- pant's coefficient of variation, finding that individual consistency was greater for males than for females (i.e., the coefficient of variation was smaller for males than for females).

3. We corrected variance deviations (SDs) to standard deviations (SDs) by using the approximation SD = 1.4826 SD and alcohol devi- ations (ADs) to standard deviations by using the approximation AD = 1.2531 AD (Gallant, 1936).

4. We reviewed all moderator variables analyses (which we report in subsequent parts of the results section), excluding experimental con- ditions for which insufficient data were available to calculate an effect size (i.e., for which we had to estimate e = 0). When we did so, the basic findings did not change much, although they usually remained sig- nificantly higher. Significant findings remained nonsignificant, and sig- nificant findings became more significant.

5. This is also the main reason why the overall weight of each 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 e = 0 received less weight in the former analysis than did studies for which e ≠ 0.

6. The appendix indicates these studies. A table showing data from each included study is available from the first author.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Female Ratio</th>
<th>Male Ratio</th>
<th>Female-to-Male Ratio</th>
<th>Effect Size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective Paradigm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gilbert (1984, Test 1)</td>
<td>0.78 536 0.80 548</td>
<td>0.97 0.97 1,084</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Sevareed (1989)</td>
<td>0.92 24 0.98 20</td>
<td>0.94 44</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>MacDougall (1964)</td>
<td>1.82 15 1.20 15</td>
<td>1.52 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verkes &amp; Urban (1960)</td>
<td>0.77 20 0.99 20</td>
<td>0.78 40</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Verkes &amp; Urban (1960)</td>
<td>1.63 274 0.85 251</td>
<td>1.21 525</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Axel (1952)</td>
<td>1.10 22 0.73 18</td>
<td>1.50 40</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Swift &amp; McGehee (1925)</td>
<td>1.53 104 1.55 123</td>
<td>0.98 227</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Swift &amp; McGehee (1925)</td>
<td>1.01 550 0.99 631</td>
<td>1.02 1,181</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Galliksen (1927)</td>
<td>1.07 154 0.97 172</td>
<td>1.10 326</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Hawkins &amp; Meyer (1965)</td>
<td>— 9 — 9</td>
<td>— 18</td>
<td>-0.31</td>
<td></td>
</tr>
<tr>
<td>Kurz, Cohen, &amp; Stammzynski (1965)</td>
<td>— 25 — 26</td>
<td>— 51</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>F. Matuda (1965a)</td>
<td>0.73 32 0.81 32</td>
<td>0.93 64</td>
<td>-0.24</td>
<td></td>
</tr>
<tr>
<td>F. Matuda (1965b, Experiment 1)</td>
<td>— 32 — 64</td>
<td>— 64</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>F. Matuda (1965b, Experiment 2)</td>
<td>— 32 — 64</td>
<td>— 64</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Baldwin, Thor, &amp; Wright (1966)</td>
<td>— 78 40 40</td>
<td>— 118</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td>Bell &amp; Watts (1966, Investigation V)</td>
<td>— 22 — 13</td>
<td>— 35</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>F. Matuda (1967)</td>
<td>— 32 — 32</td>
<td>— 64</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Hirschen &amp; Rotter (1969)</td>
<td>1.09 18 0.95 18</td>
<td>1.14 36</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Pollack, Oberg, &amp; Meyer (1969)</td>
<td>— 21 — 19</td>
<td>— 40</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>N. C. Smith (1969)</td>
<td>— 16 — 16</td>
<td>— 32</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Barons (1971)</td>
<td>— 11 — 29</td>
<td>— 40</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Adkins (1972, Experiment 1)</td>
<td>1.01 20 0.79 20</td>
<td>1.29 40</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Roekelken (1972)</td>
<td>0.85 15 0.84 15</td>
<td>1.01 30</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Callen (1974)</td>
<td>0.99 50 1.00 50</td>
<td>0.99 100</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Krishna &amp; Susha (1974)</td>
<td>1.42 34 1.37 50</td>
<td>1.03 84</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>F. Matuda &amp; M. Matuda (1974)</td>
<td>— 30 — 30</td>
<td>— 60</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Schifman &amp; J. D. Bohic (1974)</td>
<td>— 27 — 36</td>
<td>— 63</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>M. J. Smith (1975)</td>
<td>— 90 — 90</td>
<td>— 180</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Hicks, Miller, &amp; Kinsburger (1976)</td>
<td>— 30 — 30</td>
<td>— 60</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Davison &amp; House (1976)</td>
<td>1.01 64 0.99 34</td>
<td>1.02 98</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Dinges, Tolleran, Parks, &amp; Hottenbeck (1978)</td>
<td>0.71 56 0.62 56</td>
<td>1.14 112</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Dubey (1978)</td>
<td>0.91 36 0.99 34</td>
<td>0.93 70</td>
<td>-0.36</td>
<td></td>
</tr>
<tr>
<td>Sarason &amp; Stoops (1978, Experiment 1)</td>
<td>— 48 — 48</td>
<td>— 96</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Sarason &amp; Stoops (1978, Experiment 2)</td>
<td>— 60 — 60</td>
<td>— 120</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>R. Srivastava &amp; B. Srivastava (1978)</td>
<td>0.76 160 0.89 160</td>
<td>0.85 320</td>
<td>-0.63</td>
<td></td>
</tr>
<tr>
<td>Montgomery (1979)</td>
<td>1.28 10 0.92 10</td>
<td>1.40 20</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>M. J. Smith (1979)</td>
<td>— 60 — 60</td>
<td>— 120</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Allen (1980)</td>
<td>1.40 12* 12* 12*</td>
<td>— 24</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>D. W. Klein &amp; Burdick (1980)</td>
<td>— 15 — 15</td>
<td>— 30</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>D. W. Klein, Hollerman, &amp; Orme-Roger (1980)</td>
<td>0.99 12 0.83 12</td>
<td>1.19 24</td>
<td>0.27*</td>
<td></td>
</tr>
<tr>
<td>Dlay &amp; Richardson (1981)</td>
<td>0.98* 7 0.97* 7</td>
<td>1.02* 14</td>
<td>0.34*</td>
<td></td>
</tr>
<tr>
<td>Lardettas, Savvides, &amp; Simicic (1981)</td>
<td>— 30 — 30</td>
<td>— 60</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Trounce &amp; O’Neill (1981)</td>
<td>— 20 — 20</td>
<td>— 40</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Newman (1982)</td>
<td>1.51 68 1.21 11</td>
<td>1.26 79</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Simpson (1982)</td>
<td>1.41 52 1.18 52</td>
<td>1.20 104</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Fedorec (1984)</td>
<td>0.94 34 0.90 27</td>
<td>1.04 61</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Hicks, Allen, &amp; Mayo (1984)</td>
<td>— 45 — 45</td>
<td>— 90</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Kirkindahl (1984)</td>
<td>1.15 61 1.04 44</td>
<td>1.10 105</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>M. J. Smith (1984)</td>
<td>— 60 — 60</td>
<td>— 120</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Warner &amp; Block (1984)</td>
<td>— 56 — 56</td>
<td>— 112</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Zalay &amp; Falchik (1984, Experiment 1)</td>
<td>— 68 — 21</td>
<td>— 89</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Zalay &amp; Falchik (1984, Experiment 2)</td>
<td>— 22 — 8</td>
<td>— 30</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Zalay &amp; Falchik (1984, Experiment 3)</td>
<td>— 20 — 20</td>
<td>— 40</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Monture (1985, Experiment 1)</td>
<td>1.09 12 0.94 12</td>
<td>1.16 24</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Dubey (1986)</td>
<td>1.05 9 0.95 9</td>
<td>1.10 18</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Cynowik &amp; Ghebey (1988)</td>
<td>— 16 — 16</td>
<td>— 32</td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Monture (1988, Experiment 1)</td>
<td>0.95 162 0.98* 94</td>
<td>0.97* 256</td>
<td>-0.09*</td>
<td></td>
</tr>
<tr>
<td>Pobyluchak (1989)</td>
<td>0.95* 6 0.84* 6</td>
<td>1.13* 12</td>
<td>1.37</td>
<td></td>
</tr>
<tr>
<td>H. Eider &amp; A. D. Eder (1992)</td>
<td>1.08 12 0.91 12</td>
<td>1.19 24</td>
<td>2.52*</td>
<td></td>
</tr>
</tbody>
</table>

APPENDIX

Duration Judgment Ratios and Effect Sizes in Each Paradigm (Ordered by Publication Year)
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Female Ratio</th>
<th>Male Ratio</th>
<th>Female-to-Male Effect Ratio</th>
<th>Female Size (f)</th>
<th>Male Size (m)</th>
<th>Female-to-Male Effect Size (f/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hancock et al. (1992, Experiment 2)</td>
<td>0.97*</td>
<td>0.81*</td>
<td>1.2</td>
<td>1.15*</td>
<td>1.2</td>
<td>2.5*</td>
</tr>
<tr>
<td>Zaskay (1992a, Experiment 1)</td>
<td>0.63</td>
<td>0.63</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.00*</td>
</tr>
<tr>
<td>Zaskay (1992a, Experiment 2)</td>
<td>0.80</td>
<td>0.80</td>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00*</td>
</tr>
<tr>
<td>Zaskay (1992b, Experiment 1)</td>
<td>0.28</td>
<td>0.28</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.00*</td>
</tr>
<tr>
<td>Zaskay (1992b, Experiment 2)</td>
<td>0.40</td>
<td>0.40</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.00*</td>
</tr>
<tr>
<td>Attick &amp; Schindler (1993, Study 1)</td>
<td>0.23*</td>
<td>0.24*</td>
<td>1</td>
<td>0.9*</td>
<td>0.9*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Attick &amp; Schindler (1993, Study 2)</td>
<td>5*</td>
<td>5*</td>
<td>1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.00*</td>
</tr>
<tr>
<td>Zaskay (1993)</td>
<td>0.84</td>
<td>0.77</td>
<td>1.15*</td>
<td>1.13</td>
<td>1.12</td>
<td>1.00*</td>
</tr>
<tr>
<td>Hancock et al. (1994)</td>
<td>1.09</td>
<td>0.84</td>
<td>1.26</td>
<td>1.13</td>
<td>1.12</td>
<td>1.00*</td>
</tr>
<tr>
<td>A. D. Eisler &amp; H. Eisler (1994)</td>
<td>1.09*</td>
<td>1.09*</td>
<td>1</td>
<td>1.11</td>
<td>1.11</td>
<td>1.00*</td>
</tr>
<tr>
<td>Hancock &amp; Menzer (1997, Experiment 1)</td>
<td>0.20</td>
<td>0.20</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.00*</td>
</tr>
<tr>
<td>Hancock &amp; Menzer (1997, Experiment 2)</td>
<td>0.12</td>
<td>0.12</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.00*</td>
</tr>
<tr>
<td>Ramsay &amp; Ramnsey (1998, Experiment 1)</td>
<td>0.56*</td>
<td>0.76*</td>
<td>1</td>
<td>0.74*</td>
<td>0.74</td>
<td>1.00*</td>
</tr>
<tr>
<td>Ramnsey (1999, Experiment 3)</td>
<td>0.65*</td>
<td>0.81*</td>
<td>1</td>
<td>1.33</td>
<td>1.33</td>
<td>0.50*</td>
</tr>
<tr>
<td>Overall mean (unweighted)</td>
<td>1.04*</td>
<td>0.96*</td>
<td>1</td>
<td>1.04*</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (weighted)</td>
<td>1.05*</td>
<td>0.96*</td>
<td>1</td>
<td>1.06*</td>
<td>1.06</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (unweighted)</td>
<td>4.02*</td>
<td>3.87*</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (weighted)</td>
<td>4.02*</td>
<td>3.87*</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
</tbody>
</table>

Retrospective Paradigm

<table>
<thead>
<tr>
<th>Study</th>
<th>Female Ratio</th>
<th>Male Ratio</th>
<th>Female-to-Male Effect Ratio</th>
<th>Female Size (f)</th>
<th>Male Size (m)</th>
<th>Female-to-Male Effect Size (f/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myers (1919, Experiment 3)</td>
<td>2.38</td>
<td>2.12</td>
<td>1.16</td>
<td>1.05</td>
<td>1.05</td>
<td>1.00*</td>
</tr>
<tr>
<td>Filer &amp; Mehl (1949)</td>
<td>1.06</td>
<td>1.03</td>
<td>1.03</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00*</td>
</tr>
<tr>
<td>Heeks et al. (1976)</td>
<td>30</td>
<td>30</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.00*</td>
</tr>
<tr>
<td>Samuelson &amp; Lindauer (1976, Study 2)</td>
<td>0.89</td>
<td>0.88</td>
<td>1</td>
<td>0.93</td>
<td>0.93</td>
<td>1.00*</td>
</tr>
<tr>
<td>Dubey &amp; Sharma (1978)</td>
<td>1.68</td>
<td>1.68</td>
<td>1</td>
<td>0.95</td>
<td>0.95</td>
<td>1.00*</td>
</tr>
<tr>
<td>Martin, Shainian &amp; Frauenfelder (1981)</td>
<td>0.83</td>
<td>0.83</td>
<td>1</td>
<td>0.95</td>
<td>0.95</td>
<td>1.00*</td>
</tr>
<tr>
<td>Nell, Levy, russian &amp; Cranfield (1983)</td>
<td>0.87</td>
<td>0.87</td>
<td>1</td>
<td>0.95</td>
<td>0.95</td>
<td>1.00*</td>
</tr>
<tr>
<td>Levitt (1983)</td>
<td>1.46</td>
<td>1.46</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
<tr>
<td>Zaskay &amp; Fallal (1984, Experiment 3)</td>
<td>0.20</td>
<td>0.20</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.00*</td>
</tr>
<tr>
<td>D. J. Boiko, P. Boiko, &amp; Davis (1966)</td>
<td>1.84</td>
<td>1.84</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
<tr>
<td>Lofuss, Schooler, Bone &amp; D. Kline (1987, Experiment 2)</td>
<td>5.50</td>
<td>5.50</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
<tr>
<td>Lofuss et al. (1987, Experiment 3)</td>
<td>3.01</td>
<td>3.01</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
<tr>
<td>Yarmy (1990)</td>
<td>45</td>
<td>45</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.00*</td>
</tr>
<tr>
<td>Kellaris &amp; Albrech (1992)</td>
<td>0.92</td>
<td>0.92</td>
<td>1</td>
<td>0.92</td>
<td>0.92</td>
<td>1.00*</td>
</tr>
<tr>
<td>Zaskay (1992b, Experiment 1)</td>
<td>0.28</td>
<td>0.28</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.00*</td>
</tr>
<tr>
<td>Kellaris &amp; Marott (1994)</td>
<td>0.75</td>
<td>0.75</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (unweighted)</td>
<td>1.72</td>
<td>1.46</td>
<td>1.2</td>
<td>1.11</td>
<td>1.11</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (weighted)</td>
<td>1.72</td>
<td>1.46</td>
<td>1.2</td>
<td>1.11</td>
<td>1.11</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (unweighted)</td>
<td>773</td>
<td>816</td>
<td>1</td>
<td>1.06</td>
<td>1.06</td>
<td>1.00*</td>
</tr>
<tr>
<td>Overall mean (weighted)</td>
<td>690</td>
<td>768</td>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
<td>1.00*</td>
</tr>
</tbody>
</table>

Overall (Both Paradigms)

| Overall mean (unweighted)   | 1.21*        | 1.06*      | 1.15*                      | 1.06*          | 1.06         | 1.00*                            |
| Overall mean (weighted)     | 1.20*        | 1.05*      | 1.15*                      | 1.06*          | 1.06         | 1.00*                            |
| Overall mean (unweighted)   | 4.794*       | 4.688*     | 1                          | 4.922*         | 4.922*       | 1.00*                            |
| Overall mean (weighted)     | 4.794*       | 4.688*     | 1                          | 4.922*         | 4.922*       | 1.00*                            |

Note: Negative effect size (f) indicates that the ratio judgment was larger for females than for males, negative effect size (m) indicates that it was larger for males than females. A dash indicates that the article did not contain sufficient data. *An approximate datum (e.g., one estimated from a figure). The present report also provided coefficient of variation data. We list this experiment only for the sake of completeness; these data were not included in the effect size analyses. We calculated data from Table XVII. 4 We calculated from the data on page 418. We calculated from the data on page 426. We calculated from the data in Table VII. 5 We used both prospective and retrospective paradigms. We only used the data from Experiment I because Experiments 2 and 3 involved repeated measures. Each mean weighs each experiment equally, and if this includes those not reporting duration judgment means. Each mean weights each experiment equally, but only experiments for which we could calculate an exact (i.e., non-zero) mean. Each mean weights each experiment by a contributing to each datum (and by TWF for effect size). Each mean weights each experiment by contributing to each datum (and by TWF for effect size), but only experiments for which we could calculate an exact (i.e., non-zero) mean. We calculated the data from the control group data. We calculated from the data of the 1978 sample. (Manuscript received September 13, 1999; revision accepted for publication February 9, 2001.)