Developmental Review **19,** 183–211 (1999)
Article ID drev.1998.0475, available online at http://www.idealibrary.com on IDEAL®

Developmental Changes in Human Duration Judgments: A Meta-Analytic Review

Richard A. Block

Montana State University

Dan Zakay

Tel Aviv University, Israel

and

Peter A. Hancock

University of Minnesota

We reviewed 20 experiments comparing duration judgments made by children versus adolescents and adults. All used a prospective paradigm, in which participants knew they would have to make duration judgments. Meta-analyses revealed substantial age-related differences: Compared to older participants, children make larger verbal estimates, comparable productions, and shorter reproductions of duration. Children's duration judgments also show greater interindividual variability. We discuss physiological hypotheses concerning pacemaker rate and temperature or metabolic rate, along with cognitive hypotheses concerning duration units, memory processes, attentional resources, and impatience and waiting. At least two explanations are needed: Children have not yet accurately learned verbal labels for duration experiences, and they are impatient during relatively empty durations. Both can be interpreted in terms of an attentional-gate model. © 1999 Academic Press

Psychological time involves processes by which an organism adapts to and represents the temporal properties of environmental events. For more than a century, experimental psychologists have studied aspects of psycho-

This study was supported in part by grants from the United States–Israel Binational Science Foundation and the MONTS Program at Montana State University. We thank Robin Hall for helping to locate and retrieve articles; Iren Putashkin, Dida Sasson, and Fusako Matsui for helping to translate articles; and Erik Arthur for assisting with coding. We also thank Alice Eagly, Anna Eisler, Hannes Eisler, Diane Halpern, Charles A. Pierce, Grover Whitehurst, and two anonymous reviewers for commenting on preliminary versions of the manuscript. Preliminary versions of this study were presented at the meetings of the Psychonomic Society, November 1994, and the International Society for the Study of Time, July 1995.

Address correspondence and reprint requests to Richard A. Block, Department of Psychology, Montana State University, Bozeman, Montana 59717-3440. E-mail: block@montana.edu.

0273-2297/99 $30.00

logical time, such as simultaneity, successiveness, temporal order, duration, and temporal perspective (Block, 1990; Guyau, 1890/1988; James, 1890; Michon & Jackson, 1985). Some early essays focused heavily on the experience of duration. Early researchers studied duration judgments for their intrinsic interest, such as in the context of psychophysical investigations (see Woodrow, 1951). More recently, and arguably more importantly, researchers study duration judgment processes in order to clarify general physiological, memorial, and attentional processes. Organisms must time durations ranging from milliseconds to years in order to represent their external environment. However, the predominant use of temporal information concerns behavior in the specious present, which is on the order of several seconds (Iberall, 1992; Pöppel, 1985/1988). For example, crossing a busy street safely requires estimation of speed and time, or time to contact (Hancock & Manser, 1997). Because many such everyday situations involve duration estimates, to understand behavior it is important to comprehend the development of underlying processes. Studying the development of duration judgment processes may also shed light on what is finally attained—that is, on the processes that adults use in judging durations. Finally, in applied settings such as courtroom (eyewitness) testimony, it is important to know how accurate children's duration judgments may be.

Development of Duration Experience and Judgment

Piaget (1946/1969) thought that temporal cognition is gradually acquired during several developmental stages, with a child's thinking differing from an adult's until the child completes all the developmental periods of logical and abstract thinking. Piaget, who was stimulated by a question of Albert Einstein concerning the development of the notion of time, conducted a series of experiments. He showed children moving pairs of objects and asked them to indicate which object had moved for a longer duration. In some experiments, two objects started moving at the same spatial location and at the same time, moved for the same duration, but traveled different distances (because one traveled at a greater velocity). In this case, preoperational children tended to think that the object which had traveled a greater distance had also moved for a longer duration. Only later in development were children able to judge that both objects had moved for the same duration. In Piaget's view, the development of temporal cognition involves the gradual learning of coordinative relationships between spatial information (e.g., distance traveled) and temporal information. Although Piaget was more concerned with logical relations than with direct perception of duration, his view suggests that children may use different processes than those adults use in judging durations. Piaget's work was preceded by a number of previous substantive investigations, and we now consider these earlier experiments.

The first experiments comparing the magnitude of children's duration judgments and that of adolescents and young adults had been conducted much earlier, around the turn of the century (Gilbert, 1894; Seashore, 1899). In a review of some early experiments, Goldstone and Goldfarb (1966) concluded that:

> Younger children overestimate time to a greater extent than [do] older children and adults. Since this finding is independent of sense mode, and response requirements, as well as psychophysical method and other contextual or procedural factors . . . it may be considered a general statement regarding time perception by children. (p. 480)

Goldstone and Goldfarb did not offer a very satisfying explanation for these alleged differences. They mainly said that ''it became evident we were working with a biological clock with biochemical and physiological systems, and a psychological clock with sensory and conceptual systems'' (pp. 482–483). In short, various experiments have compared duration judgments made by participants of different ages. Although these authors and others (e.g., Friedman, 1978) have described developmental differences in duration judgments, no systematic understanding has emerged. Few recent reviews of experimental findings have even addressed this issue.

Models of Duration Judgment Processes

Developmental differences in duration judgments may originate in various physiological and cognitive processes. These include: (a) the rate of a person's biological processes, such as those thought to underlie an internal pacemaker or a similar component of an internal clock; (b) the person's brain temperature or basal metabolic rate; (c) the way a person translates between objective and subjective units of duration; (d) the person's memory processes, such as those that mediate the encoding and forgetting of information; (e) the person's allocation of attentional resources, especially attention to time; and (f) the person's impatience and tolerance of a situation requiring waiting or delaying a response. Because the methods used to investigate developmental differences in duration judgments may be sensitive to some or all of these physiological and cognitive processes, we need to consider models and the specific predictions they make in different temporal judgment situations.

Theorists have proposed various models to explain duration judgment processes (Block, 1990). One approach emphasizes physiological elements. For example, an internal clock, consisting of a biological pacemaker and other components, may subserve time-related behavior. Influences such as brain temperature and metabolism, psychoactive drugs, and arousal level may affect the pacemaker rate (see Hancock, 1993). Another kind of model is that duration is a cognitive construction which is influenced mainly by attention and memory processes. Although the two views may be construed as separate proposals, some researchers have proposed hybrid models (e.g., Treisman, 1963; Zakay & Block, 1996, 1997). Zakay and Block's attentional-gate model, for example, proposes that duration judgments depend on both physiological elements (e.g., arousal level, which influences the rate of a pacemaker) and cognitive elements (e.g., an attentional gate, which influences how much temporal information is transmitted to a memory store). Children and adults may make quantitatively different duration judgments if any of the components of the model function differently with age. Several methodological factors may differentially emphasize the role of various components; we now discuss two of the more important ones.

Duration judgment paradigm. If a person is aware during a time period that its duration needs to be estimated (the defining characteristic of what is called the *prospective paradigm,* or *experienced duration*), duration judgments are directly related to the amount of attention a person allocates to temporal information. If a person is not aware of this until after the time period (the defining characteristic of what is called the *retrospective paradigm,* or *remembered duration*), duration judgments are directly related to encoded and retrieved memory information, such as concerning the remembered number of events or amount of change in cognitive context. All of the experiments included in the present meta-analyses used the prospective paradigm, so our theoretical interpretations focus on explanations for duration judgments made under that paradigm only. However, it is clear that important theoretical and practical information may be obtained by comparing children's and adults' retrospective duration judgments, and we encourage such an endeavor by noting the dearth of such studies.

Duration judgment method. Several commonly used methods involve comparing a duration experience with internal (often called *reference memory*) information concerning labels for duration units, such as seconds and minutes. In the method of verbal estimation, a person uses such a numerical label to judge a past duration (e.g., ''estimate the length of that duration in seconds''). In the method of production, a person delimits an objectively measured duration corresponding to a subjectively defined time duration (e.g., ''say *start,* then say *stop* when it seems like 60 s has elapsed''). In the method of repeated production, a person delimits consecutive objectively measured durations of a requested length, usually 1 s (e.g., ''press this button every 1 s until I tell you to stop''). These methods are suitable to investigate developmental differences in the translation between experienced duration and conventional duration units. They may also reveal effects of variables that may influence the rate of internal processes, such as a biological pacemaker, so long as a person has not yet learned to recalibrate—that is, to adjust sufficiently the translation for the effect of the variables. Researchers have used these methods to study acute manipulations thought to influence the rate of internal timekeeping processes, such as drugs (e.g., Frankenhaeuser, 1959; Hicks, 1992).

In the reproduction method, a person experiences a target duration of a certain objective length, then operatively delimits a second duration to estimate the target duration. This method does not require the knowledge or use of conventional duration units. However, reproductions may be ''an index only of the consistency of the subjective time base; [they provide] no information as to the rate itself'' (Cahoon, 1969, p. 261). Even if the rate of physiological and cognitive processes varies with age, the same rate ordinarily subserves a person's experiencing the target duration and reproducing it (cf. Bindra & Waksberg, 1956). Thus, the reproduction method, which is very useful in some experimental contexts, may not reveal much about developmental differences in duration judgments. However, it may detect developmental differences if it is used in the framework of psychophysical studies, in which duration is varied. If relatively fewer subjective temporal units (of whatever basis) are stored as a time period lengthens, then the exponent of the psychophysical function may be less than 1. In one study of a person suffering from anterograde amnesia, reproductions of longer durations were abnormally short (Richards, 1973). Hence, the reproduction method may be sensitive to forgetting of information from the target duration. In addition, both the production method and the reproduction method may have confounding from extraneous variables (e.g., desire to terminate the experiment sooner, impatience, or inability to delay a response).

Droit-Volet (1998) reviewed extant literature and reported new data on time production in very young children (3-year-olds and 5½-year-olds). She found that the younger children's productions were not influenced much by temporal instructions, but they were by nontemporal (in this case, response force) instructions. Pouthas (1993) argued that most young children (below the age of about 7 years) have not yet learned how to use conventional duration units (seconds, minutes, and so on) in a reasonably accurate way. If this learning occurs at different rates, then children's duration judgments should show greater intraindividual and interindividual variability than adults' duration judgments do. If various children tend systematically to err in the same direction in this translation between subjective and objective units, they may make either smaller or larger verbal estimates of duration than do adults and either longer or shorter productions of duration than do adults. If children and adults differ only in the translation between objective and subjective units, however, their reproductions of duration may not differ, because reproductions do not require knowledge of verbal units or a fully developed translation process. Finally, children may appear to be more or less sensitive than are adolescents and adults to the stimulus content presented or the processing task required during a particular duration. This may be the case because they have not yet learned how various factors influence duration experience. Adolescents and adults may verbally estimate and produce durations in a way that compensates for the effect of the content or the processing task on their subjective experience (Arons & London, 1969).

Specific Goals of the Present Meta-Analyses

In the present review, we quantitatively evaluated evidence on developmental differences in duration judgments. We first investigated the question of whether children show greater, lesser, or comparable ratios of subjective duration to objective duration than do adolescents and young adults. Then we determined whether or not any variables (such as duration judgment method) moderated duration judgment magnitude. The finding of significant moderator variables clarifies theories concerning processes that may show developmental differences. (The finding of a moderator variable is analogous to the finding of an interaction effect in primary research.) Because most experimental reports also included appropriate information, we also accumulated primary-level statistics based on a common scale of measurement, the ratio of subjective to objective duration. In this regard, the present review differs from most meta-analytic reviews: Accumulating primary-level statistics across conditions and experiments clarifies meta-analytic statistics (Block & Pierce, 1998).

Many of the reviewed articles reported data on the interindividual variability of duration judgments, such as standard deviations. (None reported any data on intraindividual variability, another potentially interesting measure.) In spite of this, no researcher used inferential statistics to test hypotheses on intraindividual or interindividual variability, and surprisingly few commented on such differences in variability (but see Goldstone & Goldfarb, 1966). Because interindividual variability is theoretically relevant, we quantitatively reviewed data on it. The present meta-analysis is the first to report on interindividual variability in duration judgments. Analyses of interindividual variability of duration judgments may provide insights concerning whether participants of various ages use similar or different kinds of processes. Greater interindividual variability is expected if different participants use different processes, which affect duration judgments differently. It may also be expected if different participants use the same (or similar) processes, but the intraindividual variability of those processes differs (especially if only a few duration judgments are obtained).

Finally, we discuss one experimental report that contained data on the exponent (β) of the psychophysical function relating subjective and objective duration. Differences in the exponent reveal whether age and duration length interact to influence duration judgments.

METHOD

Sample of Studies

We searched a database containing more than 9000 references on the psychology of time (Block & H. Eisler, 1998), which includes articles from the following sources: *Psychological Abstracts* (1923–1966) and *PsycINFO* (1967–1997), using the keywords *time perception* and *time estimation;* and

Medline (1966–1997), using the keyword *time perception.*[1] The database includes references from published bibliographies, articles, books, book chapters, and our files (see Block & Zakay, 1997, for details). We also searched *Social Sciences Citation Index* (Social SciSearch, 1977–1997) for articles that cited already retrieved articles (e.g., LeBlanc, 1966, 1969), and we checked the reference lists of all included articles.

To be included in the present meta-analyses, an experimenter must have studied normal human participants of diverse ages, and a published article must have contained quantitative data on duration judgment magnitude in children and either adolescents, young adults, or both. We compared children to adolescents and young adults, where older participants' age was a coded (potential moderator) variable. This process excluded experiments in which age was studied only within a single category, such as only children (e.g., Arlin, 1986a, 1986b). For a few experiments in which participants' ages slightly overlapped two categories (e.g., adolescents and young adults), we used the predominant category. The process excluded experiments in which researchers did not obtain and report duration judgment magnitude, such as: (a) experiments assessing time-of-arrival judgments, tau effects, and kappa effects, all of which involve a substantial spatial component (e.g., Matsuda, 1974, 1989; Matsuda, Miyazaki, & Matsuda, 1983); (b) experiments in which researchers reported only data on duration judgment error, percentage correct judgments, or percentage underestimation or overestimation (e.g., Dmitriyev, 1980; Elkine, 1928; Gilliland & Humphreys, 1943; Herman, Norton, & Roth, 1983; Tejmar, 1962); (c) experiments in which participants made qualitative (e.g., same/different and shorter/longer) judgments, such as duration-discrimination judgments, with only discrimination threshold data reported[2] (e.g., Goldstone & Goldfarb, 1966; Smythe, 1956; Smythe & Goldstone, 1957); (d) experiments with sample sizes less than five (e.g., Richards, 1964); (e) experiments in which participants received feedback, thereby learning to improve their duration judgments (e.g., Fraisse & Orsini, 1958; Matsuda & Matsuda, 1987); and (f) experiments that we would have included except that researchers did not report sufficient statistics (e.g., mean judgment, numbers of participants, or inferential statistics) to estimate an effect size or duration judgment ratio (e.g., Dmitriyev, 1980; Wallon, Evart-Chmielniski, & Denjean-Raban, 1957).

We did not want to distort the analyses by including highly variable data from very young children, many of whom do not yet grasp the concept of duration or know how to translate between subjective and objective duration units (Block, 1990; Friedman, 1978; Pouthas, 1993). Thus, we excluded data

[1] We included any relevant study listed in *PsycINFO* or *Medline* as of December, 1997.

[2] Without assuming a mathematical model, this kind of judgment cannot be compared to quantitative duration judgments of the kind meta-analyzed here. Including them would subject the meta-analysis to an ''apples and oranges'' criticism (Sharpe, 1997).

from children younger than 7.0 years of age (e.g., A. D. Eisler, H. Eisler, Guirao, & Harris, 1995; Landaeta, Saavedra, & Simicic, 1981). Several articles contained statistics on subgroups within a given age class; for these, we pooled data and effect-size estimates from the respective subgroups.

Coded Variables

Based on a consideration of theoretically relevant variables, we coded the following variables from each experiment and from each within-experiment condition: (a) publication year; (b) participants' sex (female, male, both, or unknown); (c) participants' age (child [7.0–12.9 years of age], adolescent [13.0–17.9 years of age], and young adult, such as most samples involving college students [18.0–29.9 years of age]); (d) modality (visual, auditory, tactile, or mixed) of stimuli presented during the duration, if any; (e) duration length predominantly used (very short [4.9 s or shorter], short [5.0–14.9 s], moderate [15.0–59.9 s], or long [60.0 s or longer]); (f) duration judgment method (verbal estimation, production, repeated production, or reproduction); and (g) total number of duration judgments made by each participant during the experiment. Two authors coded all study attributes independently, resolving any disagreements by discussion.

We also coded several other variables, but for these only one class of the variable was adequately represented in the meta-analyses; the others contained fewer than three effect-size estimates. No conclusions are possible for such variables. Future research may reveal whether these variables moderate age effects on duration judgments. All (or nearly all) experiments: (a) used a prospective duration judgment paradigm; (b) did not state whether participants' watches were removed; (c) presented either no stimuli or one simple, continuous stimulus; (d) did not segment the duration with salient markers or high-priority events; (e) had few environmental or background changes; (f) required only passive/covert and easy/shallow processing of any presented stimuli; (g) did not suggest or require the use of chronometric counting; (h) did not interpose any delay, except possibly for brief instructions, preceding the duration judgment; (i) did not require memory for any presented information; (j) had no changes in type or level of processing; and (k) had no concurrent task.

Effect-Size Analyses

Two authors independently estimated effect sizes, resolving disagreements by discussion. Each effect size was calculated as g, the difference between the mean duration judgment given by participants of two age classes divided by the pooled standard deviation (Hedges & Olkin, 1985), using the computer software DSTAT (Johnson, 1989, 1993). If the researcher only reported a nonsignificant finding, with no inferential statistic, we assumed that $g = 0$ if we could not determine the direction of the effect (from either means or a verbal description). Effect sizes were calculated separately, whenever possible, for different levels of manipulated variables (e.g., for different duration lengths). To provide a single measure for each experiment, we averaged all such separately calculated effect sizes. Each *g* was converted to *d* by correcting it for bias, weighting by the reciprocal of its variance (Hedges, 1981; Hedges & Olkin, 1985). Then we combined the *d*s by separately calculating unweighted and weighted means.

If a researcher manipulated a potential moderator variable (e.g., used different duration judgment methods) and provided adequate statistics for us to calculate separate effect-size estimates for each level of the variable, we did so for that moderator analysis. Thus, each moderator analysis contained mainly experiment effect sizes, but also a few within-experiment effect sizes. Using more than one effect-size estimate from the same experiment violates the assumption that effect sizes are independent. However, this kind of violation does not substantially affect statistical precision (Tracz, 1984/1985; Tracz, Elmore, & Pohlmann, 1992). If we had not used more than one effect-size estimate from experiments that manipulated a potential moderator variable, we could not have properly conducted the moderator analyses, because we would have had to discard (or code into a *mixed* category) some of the most relevant information.[3]

The homogeneity of each set of *d*s was tested to determine whether the conditions shared a common effect size. If there was significant heterogeneity of effect sizes, as indicated by the statistic *Q*, we attempted to account for it with coded or manipulated study attributes. Two coded variables, publication year and total number of duration judgments, are continuous. We tested those variables by using a weighted least-squares regression model (Hedges, 1982b; Hedges & Olkin, 1985), using SPSS and DSTAT (see Johnson, 1989). We tested all other coded variables by using categorical models (Hedges, 1982a; Hedges & Olkin, 1985), as implemented by DSTAT. (In a few cases, we combined two similar classes of a variable if there were fewer than three effect-size estimates in a given class.) These techniques yielded a between-classes effect, revealing whether that variable is a significant moderator of the age effect (i.e., whether one would expect to find an interaction effect in an experiment). The mean effect size (d_{i+}) was calculated for each category, with each effect size weighted by the reciprocal of its variance. When appropriate, we used DSTAT to perform post hoc paired contrasts of d_{i+} in each pair of classes. All *p* values for these simple contrasts are two-tailed.

The order in which coded attributes (potential moderator variables) are listed reflects our judgment about the relative importance (from most to least) of each variable. This judgment was based on several criteria: (a) the size and significance of the between-classes effect (Q_B) in the relevant categorical

[3] When we used only one randomly selected effect-size estimate from each experiment, none of the results of the moderator analyses changed in any substantial way.

model, (b) the completeness of the categorical model as indicated by each within-class homogeneity of variance (Q_{wi}), and (c) differences between primary-level statistics (see next paragraph). We also calculated correlations among coded variables in order to determine the extent to which they were relatively independent of each other.

Primary-Level Statistics

For each condition analyzed, we calculated the ratio of subjective to objective duration—hereafter called the *duration judgment ratio*—separately for each age class. This is a standard measure which many researchers calculate and report (Hornstein & Rotter, 1969). For the method of reproduction, this is the ratio of the person's reproduced (subjective) duration to the previously presented (objective) duration. For the method of verbal estimation, this is the ratio of the person's numerical (subjective) estimate to the previously presented (objective) duration. For the method of production (and repeated production), 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 commonly found negative correlation between estimates obtained by using the production (or repeated production) and verbal estimation methods. Thus, the ratio assesses the moderating influence of duration judgment method apart from the otherwise negative correlation. Using it also enables a comparison across conditions and experiments that entailed judgments of different duration lengths. We also calculated the mean ratio of younger/older duration judgments—hereafter called the *age ratio.* Two-tailed *t* tests were performed on these unweighted primary-level statistics.[4] Our description of results takes into account whether any particular comparison was significant (at $p < .05$).

We also analyzed separately the experimental reports that contained sufficient information, such as standard deviations or standard errors, to determine the relative interindividual variability of duration judgments made by participants of different ages. One cannot simply compare standard deviations when a ratio scale of measurement is involved, because standard deviations typically increase with increasing mean judgment. We instead calculated the coefficient of variation, a common psychometric measure, which is the standard deviation divided by the mean judgment. Because no article reviewed here contained coefficient of variation data, the present meta-analysis reveals new information about interindividual variability in duration judg-

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

ments. We used the program COEFVAR (Gilpin, 1993) to calculate a χ^2 value for the difference between coefficients of variation with the Bennett-Shafer-Sullivan likelihood-ratio test, then converted each χ^2 to d by using DSTAT. We also accumulated and tested primary-level statistics on coefficients of variation.

RESULTS AND DISCUSSION

A total of 20 experiments, published in 18 separate journal articles, met all criteria for inclusion. Researchers from the United States wrote 6 articles (all in English), researchers from Japan wrote 6 (3 in Japanese and 3 in English), researchers from the former Soviet Union wrote 5 (all in Russian), and a researcher from France wrote 1 (in French). The median publication year was 1968. The mean age of participants was about 9.7 years for children, 15.8 for adolescents, and 20.2 for young adults.

Duration Judgment Magnitude

Effect Sizes and Primary Statistics

All 20 experiments contributed an effect size estimate for duration judgment magnitude (see Table 1), 19 involving a between-subjects design and 1 involving a within-subjects (longitudinal) design. A total of 15 effect size estimates were calculated from means and *standard deviations,* standard errors, quartile deviations, average deviations, or mean variations[5]; 2 from means and an F value; 1 from an F value; 1 from means and a pooled standard deviation estimated from a related F value; and 1 from a reported nonsignificant effect (for which we assumed that $g = 0$, because no direction of any possible effect could be inferred[6]). We defined an effect as being positive if the duration judgment ratio (the ratio of subjective to objective duration) was greater for children and as being negative if it was greater for adolescents or young adults. The resulting weighted mean effect size $d_+ = -0.20$, 95% confidence interval (CI) $= -0.28$ to -0.13, indicating a significantly greater duration judgment ratio for adolescents and young adults than for children, $p < .001$. The d_+ of -0.20 is considered small in magnitude (Cohen, 1977). In addition, the unweighted mean duration judgment ratio for children was not significantly different from that of adolescents or young adults, $t(19) = 1.18$, $p = .25$, and the mean age ratio (1.07) was not significantly different from 1.00, $t(19) = 1.09$, $p = .29$. However, both the

[5] We converted quartile deviations (QDs) to standard deviations (SDs) according to the approximation $SD = 1.4826\ QD$, and average deviations (ADs) to SDs according to the approximation $SD = 1.2533\ AD$ (Guilford, 1936).

[6] Meta-analysts disagree on how to treat such missing data. Some recommend exclusion, some recommend replacement with the mean effect size, and some recommend other strategies (Pigott, 1994). Regardless of which strategy we used, neither the overall analysis nor any of the moderator analyses changed in any substantial way.

TABLE 1
Duration Judgment Ratios and Effect Sizes

Study	Yng ratio	Yng *n*	Old ratio	Old *n*	Yng/Old ratio	Total *N*	Effect size (*d*)
Gilbert (1894, Test 11)[a,b]	0.75	591	0.81	493	0.93	1084	−0.62
Seashore (1899)[b,c]	1.14*	97	1.19*	52	0.95*	149	−0.14*
Axel (1924)[b]	1.59	405	1.03	260	1.54	665	0.27
Fraisse (1948)[d]	0.83	24	0.92	12	0.90	36	−0.36
Goldstone et al. (1958)[b,d,e]	0.95	100	0.98	70	0.97	170	−0.08
Matsuda (1965a)[d]	0.75	48	0.88	16	0.85	64	−0.66
Matsuda (1965b, Exp. 1)[b,d]	0.86*	32	1.01*	32	0.85*	64	−1.13*
Matsuda (1965b, Exp. 2)[b,d]	0.86*	32	1.01*	32	0.85*	64	−1.11*
Dmitriyev & Tushnova (1967)[b]	1.05	26	1.01	38	1.04	64	0.17
Matsuda (1967)[d,f]	0.89*	48	0.97*	16	0.92*	64	−0.41*
LeBlanc (1969)[b,d]	1.77	35	1.34	71	1.32	106	0.68*
Dmitriyev & Voitiukova (1973)[b]	1.02	20	0.99	10	1.03	30	0.25
Matsuda & Matsuda (1974)[d]	1.04*	40	1.00*	20	1.04*	36	0.12
Matsuda & Matsuda (1976, Exp. 1)[d]	0.88	40	1.08	20	0.82	60	−0.89
Matsuda & Matsuda (1976, Exp. 2)[d]	0.82	40	1.08	20	0.76	60	−1.13
Gareyev (1977)[b,g]	2.49	20	1.32	18	1.88	38	0.42

Gareyev & Osipova (1980)[b,h]	1.51	19	0.99	22	1.53	41	0.53
Orihara (1980)[d]	0.95	30	0.87	30	1.08	60	0.36
Fedotchev (1984)[b,i]	0.99	61	0.86	61	1.15	61	0.31
Hicks et al. (1984)[b,d]	1.23*	54	1.23*	18	1.00*	72	0.00*
Overall mean (unweighted)[j]	1.12*	20	1.03*	20	1.07*	19	−0.17*
Overall mean (weighted)[k]	1.08*	1746	0.96*	1303	1.10*	3049	−0.20*

Note. Positive effect size (d) indicates that the mean duration judgment ratio (subjective/objective duration) was relatively greater for children; negative effect size (d) indicates that it was greater for older participants. Old, older participants (i.e., adolescents or young adults); Yng, younger participants (i.e., children).

* An approximate datum (e.g., one estimated from a figure).

[a] Although the authors reported absolute error in reproductions, they said the reproductions were always too short.

[b] Study compared children and adolesents.

[c] Data from Table XVIII, excluding 6-year-old children.

[d] Study compared children and young adults.

[e] We excluded the previously reported (Goldstone, Lhamon, & Boardman, 1957; Smythe & Goldstone, 1957) ''passive extimation'' data.

[f] We included only conditions under which the reproduced duration contained no stimuli.

[g] We included only data from 7–9-year-old children and 14–15-year-old adolescents; data from ''logical problem'' and young adults were excluded because of variable or different duration length.

[h] We excluded data from young adults because of variable duration length and data from 7–8-year-old children because of probable errors.

[i] Study used a within-subjects (longitudinal) design.

[j] Each mean was calculated by weighting each experiment equally.

[k] Each mean was calculated by weighting by n contributing to each datum (or by TW for effect size).

small effect size and the nonsignificant primary-level statistics must be understood in the light of moderator variables.

Moderator Variables

The homogeneity statistic indicated that effect sizes were not homogenous, $Q(19) = 159.7$, $p < .0001$, and coded study attributes were investigated as potential moderator variables to account for heterogeneity of effect sizes. Table 2 shows the results of model testing involving the only categorical variables that were significant moderators.

Duration judgment method. Because only one experiment used the method of repeated production, we combined it with the three that used the method of production. Duration judgment method significantly moderated effect sizes. For verbal estimation, the age effect was positive: Children gave larger verbal estimates than did adolescents or young adults. For the method of production, the age effect was not significant. For the method of reproduction, the age effect was negative: Children gave shorter reproductions than did older participants.[7] Simple contrasts showed that the weighted mean effect size was greatest for the method of verbal estimation, intermediate for the method of production, and least for the method of reproduction, all $p < .05$. Because all classes showed significant heterogeneity of variance, duration judgment method was not the only moderator of the age effect.

Figure 1 shows the mean duration judgment ratio as a function of the mean participants' age (in each age class) separately for each duration judgment method. Compared to older participants, children make larger verbal estimates, comparable productions, and shorter reproductions. Thus, there were significant age-related effects, but they were in opposite directions for different duration judgment methods. The overall effect size was small as a result of combining experiments that used different duration judgment methods, and it was negative because most of the experimental conditions used the method of reproduction. As shown in Fig. 1, children's mean duration judgment ratio was significantly greater for verbal estimation than for both production and reproduction; however, the latter did not differ significantly. Both adolescents and young adults gave comparable duration judgment ratios for verbal estimation, production, and reproduction. Thus, the age ratio was significantly greater for verbal estimation (1.73) than for both production (1.04) and reproduction (0.95), which did not differ.

Duration length. Across all experiments, the target duration ranged from 2.0 s to 100.0 s. Only two conditions used very short durations and only two conditions used long durations, so for the moderator analysis we combined those with short and moderate durations, respectively. Duration length mod-

[7] Additional evidence supporting this developmental trend comes from Arlin's (1986a, 1986b) studies. He found that children made reproductions shorter than the target duration, and younger children made shorter reproductions than did older children.

TABLE 2
Tests of Categorical Models for Duration Judgment Comparisons

Variable and class	Between-classes effect (Q_B)	k	Mean effect size (d_{i+})	95% *CI* for d_{i+}		Homogeneity within class (Q_{wi})[a]
				Lower	Upper	
Duration judgment method	88.26***					
Verbal estimation		3	0.33	0.17	0.48	8.65*
Production or repeated production		4	0.10	−0.09	0.29	10.18*
Reproduction		15	−0.46	−0.55	−0.37	66.39***
Duration length	93.57**					
Very short or short		12	−0.49	−0.59	−0.40	61.19***
Moderate or long		14	0.19	0.09	0.29	39.72***
Stimulus modality[b]	86.17***					
Visual		4	0.27	0.13	0.40	1.72
Auditory		7	−0.55	−0.66	−0.44	41.96***
Participants' sex[b]	21.36***					
Female		4	−0.64	−0.80	−0.48	26.98***
Male		5	−0.18	−0.29	−0.07	76.16***

Note. Positive effect sizes indicate a greater duration judgment ratio by children; negative effect sizes indicate a greater ratio by adolescents or young adults.

[a] Significance indicates rejection of the hypotheses of homogeneity.

[b] We excluded the *mixed* category from this analysis.

* $p < .05$; ** $p < .01$; *** $p < .001$.

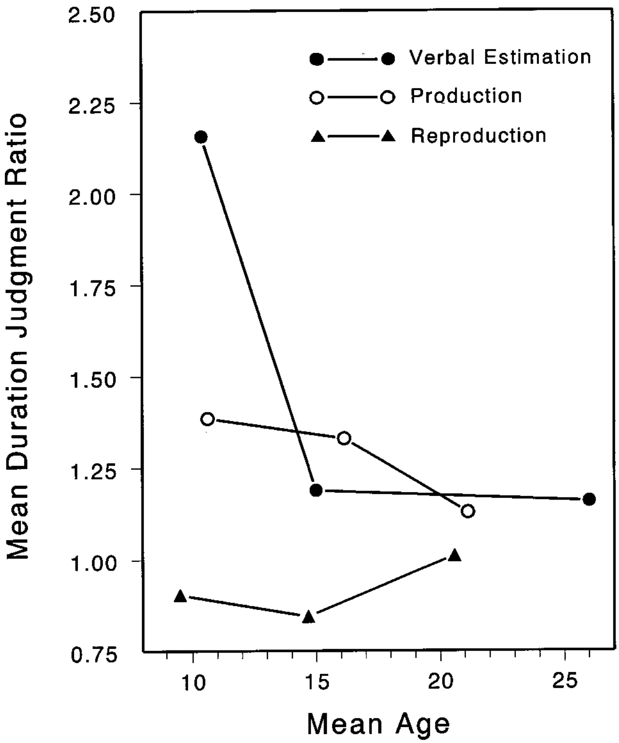

FIG. 1. Mean duration judgment ratio as a function of age class, shown separately for each duration judgment method. The mean duration judgment ratio for verbal estimates made by young adults is from Block et al. (1998).

erated effect sizes.[8] Conditions using very short or short durations showed a negative effect, whereas those using moderate or long durations showed a positive effect. Because there was significant heterogeneity of variance in both classes, duration length was not the only moderator variable. Children gave a smaller duration judgment ratio in conditions using very short or short (0.91) than moderate or long durations (1.25). Adolescents and young adults gave a comparable duration judgment ratio for very short or short (0.97) and moderate or long durations (1.04). Thus, the significantly smaller age ratio for very short or short (0.95) than for moderate or long durations (1.17) is mainly attributable to the relatively larger duration judgment ratios that children gave for long durations.

[8] Duration length was also tested as a continuous moderator variable, using the mean target duration when necessary. The linear moderating influence of duration length was significant, $Q_R = 25.07$, $p < .001$, but the model was not very well specified, $Q_E = 100.08$, $p < .001$.

Although this is an interesting finding, we note that conditions involving shorter durations all used the method of reproduction, whereas those involving longer durations tended to use the methods of verbal estimation or production. Thus, some of the moderating effect of duration length is attributable to duration judgment method instead of duration length per se. However, when the moderating influence of duration length was analyzed only for conditions using the method of reproduction, duration length was still a significant moderator, with very short or short durations showing a negative effect (i.e., children made relatively shorter reproductions) and moderate or long durations showing little or no effect.

Stimulus modality. Stimulus modality also moderated effect sizes. Conditions that used visual stimuli showed a positive effect size, whereas conditions that used auditory stimuli showed a negative effect size. However, this seems to be attributable to its correlation with duration judgment method, not anything about modality per se: Experimental conditions using the reproduction method tended to use auditory stimuli. In addition, children's duration judgment ratios did not differ significantly between conditions using visual (1.27) and auditory stimuli (0.98). Adolescents and young adults also gave comparable ratios in conditions using visual (1.04) and auditory stimuli (1.00). The age ratio also did not differ significantly between conditions using visual (1.23) and those using auditory (0.95) stimuli.

Participants' sex. Participants' sex moderated effect sizes, with females showing a significantly more negative effect size than did males. However, children's ratios did not differ between females (0.93) and males (1.06). Adolescents and young adults also showed comparable ratios for females (0.94) and males (0.97). The age ratio did not differ significantly between females (0.99) and males (1.08). Because there was significant heterogeneity of variance in both sex classes, sex was not the only moderator variable, and because the duration judgment ratios did not differ significantly, participants' sex does not appear to be a crucial moderator variable. We also note that the effect of sex also appears to be attributable to method; when we looked only at reproduction data, there was no sex difference.

Number of duration judgments. The median number of duration judgments made by each participant in an experiment was 16. The linear moderating influence of number of duration judgments was significant, $Q_R = 22.3$, $p < .0001$, but the regression model did not provide a good fit, $Q_E = 144.4$, $p < .0001$. The correlation between number of duration judgments and effect size was significantly negative, $r(19) = -.61$, $p = .004$. Thus, as the total number of duration judgments increased, the age effect decreased. The negative correlation is probably attributable to the fact that conditions involving the method of reproduction tended to entail more judgments (20.4) and to yield a more negative effect size than did conditions involving the methods of verbal estimation (4.50) and production (4.75).

Other variables. The linear moderating influence of publication year was significant, $Q_R = 27.1$, $p < .0001$, but the linear regression model did not provide a good fit, $Q_E = 27.1$, $p < .0001$. In addition, the correlation between publication year and effect size was not significant, $r(19) = .15$, $p = .52$. Older participants' age (i.e., adolescents vs. young adults) did not significantly moderate effect sizes ($Q_B = 1.96$, $p = .16$). (See Table 1 for information about the age classes of participants in each experiment.) No other variable was sufficiently represented across experiments or frequently manipulated in experiments.

Coefficient of Variation

Each of 13 experiments provided sufficient information to calculate a separate coefficient of variation for children and adolescents or young adults (see Table 3). The sign of each effect size was positive if the coefficient of variation was greater for children and negative if it was greater for adolescents or young adults. The resulting $d_+ = 0.33$, 95% $CI = 0.25$ to 0.40, indicating a greater coefficient of variation for children than for adolescents or young adults, $p < .0001$. The d_+ of 0.33 is considered small to moderate in magnitude. The unweighted mean coefficient of variation was significantly greater for children than for adolescents or young adults, $t(12) = 2.87$, $p = .014$, and the mean age ratio (1.47) was greater than 1.00, $t(12) = 14.1$, $p < .001$. Even though the effect size was small to moderate, the age ratio revealed a large difference: The mean coefficient of variation was 47% greater for children than for adolescents or young adults.

Homogeneity of effect sizes was indicated, $Q(12) = 20.4$, $p = .06$. Nevertheless, we used coded study attributes to determine whether any sufficiently represented variable accounted for heterogeneity of effect sizes. Most comparisons were between children and adolescents. Four variables were sufficiently represented: publication year, number of duration judgments, duration length, and duration judgment method.

Only duration judgment method moderated coefficient of variation effect sizes, $Q_B(2) = 9.19$, $p = .01$. Although all effect sizes were significantly greater than 0, they were greater in the eight conditions that used the method of reproduction ($d_{i+} = 0.51$) than in the three that used the method of verbal estimation ($d_{i+} = 0.29$) and in the four that used the method of production or repeated production ($d_{i+} = 0.25$). Figure 2 shows the mean coefficient of variation as a function of the mean participants' age (in each age class) separately for each duration judgment method. The mean age ratio was greater in conditions that used the method of reproduction (1.57) than the method of production (1.12), with the method of verbal estimation (1.37) intermediate.

Psychophysical Slope

Only one report (Hicks, Allen, & Mayo, 1984) contained data on the slope of the psychophysical function relating subjective duration to target duration

TABLE 3
Coefficients of Variation and Effect Sizes

Study	Yng *CV*	Yng *n*	Old *CV*	Old *n*	Yng/Old *CV*	Total *N*	Effect size (*d*)
Gilbert (1894, Test 11)[a]	0.14	591	0.11	493	1.26	1084	0.22
Seashore (1899)[a,b]	0.42*	97	0.31*	52	1.48*	149	0.48
Axel (1924)[a]	1.42	405	0.91	260	1.56	665	0.32
Fraisse (1948)[c]	0.31	12	0.31	24	0.99	36	0.02
Goldstone et al. (1958)[a,c,d]	0.40	100	0.24	70	1.63	170	0.57
Matsuda (1965a)[c]	0.29	48	0.13	16	2.19	64	0.92
Dmitriyev & Tushnova (1967)[a]	0.24	26	0.16	38	1.54	64	0.61
LeBlanc (1969)[a,c]	0.42	35	0.39	71	1.08	106	0.12
Dmitriyev & Voitiukova (1973)[a]	0.17	20	0.09	10	1.82	30	0.74
Gareyev (1977)[a,e]	1.00	20	0.61	18	1.71	38	0.48
Gareyev & Osipova (1980)[a,f]	0.46	19	0.57	22	0.81	41	−0.22
Orihara (1980)[c]	0.21	30	0.16	30	1.29	60	0.66
Fedotchev (1984)[a,g]	0.50	61	0.32	61	1.59	61	0.56
Overall mean (unweighted)[h]	0.46*	13	0.33*	13	1.47*	13	0.32
Overall mean (weighted)[i]	0.58*	1476	0.36*	1153	1.41*	2568	0.33

Note. Positive effect size (*d*) indicates that the mean coefficient of variation (*CV*) was larger for children; negative effect size indicates that it was larger for older participants. Old, older participants (i.e., adolescents or young adults); Yng, younger participants (i.e., children).

* An approximate datum (e.g., one estimated from a figure).

[a] Study compared children and adolescents.

[b] Data from Table XVIII, excluding 6-year-old children.

[c] Study compared children and young adults.

[d] We excluded the previously reported (Goldstone et al., 1957; Smythe & Goldstone, 1957) ''passive estimation'' data.

[e] We included only data from 7–9-year-old children and 14–15-year-old adolescents; data from ''logical problem'' and young adults were excluded because of variable or different duration length.

[f] We excluded data from young adults because of variable duration length and data from 7–8-year-old children because of probable errors.

[g] Study used a within-subjects (longitudinal) design.

[h] Each mean was calculated by weighting each experiment equally.

[i] Each mean was calculated by weighting by *n* contributing to each datum (or by *TW* for effect size).

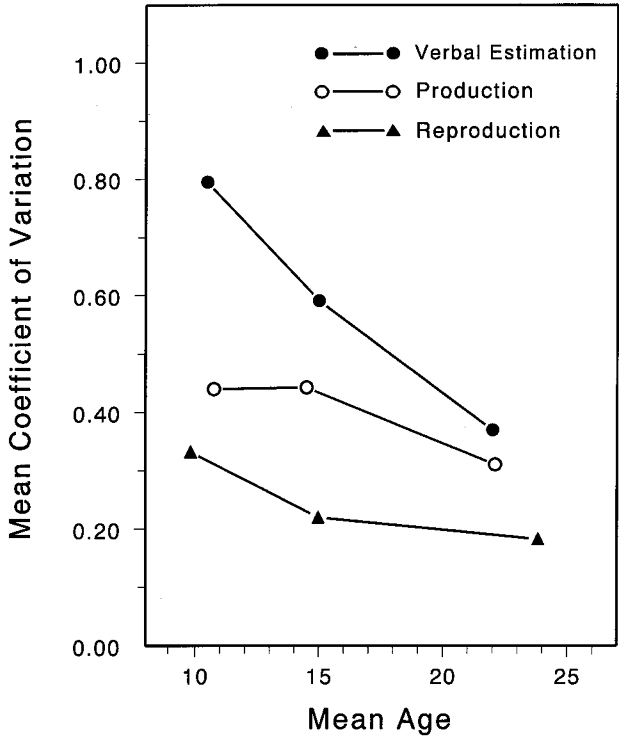

FIG. 2. Mean coefficient of variation as a function of age class, shown separately for each duration judgment method. The mean coefficient of variation for verbal estimates made by young adults is from Block et al. (1998).

in children compared to adolescents or young adults. Hicks et al. found that children and young adults did not show significantly different slopes. Data are clearly too scarce to draw any defensible conclusion.

DISCUSSION

The meta-analyses reveal several important findings regarding the influence of age on duration judgments. Considering duration judgment magnitude, the overall weighted mean effect size was small and negative. When we tested categorical models investigating potential moderator variables and aggregated primary-level statistics (duration judgment ratios), duration judgment method was revealed to be the most important variable moderating heterogeneity of effect sizes across experimental conditions. Children differ from adolescents and young adults in that they make larger verbal estimates, comparable productions, and shorter reproductions of duration. (The overall negative effect size is a consequence of the fact that 15 of the 22 conditions used the method of reproduction.) Several other variables also moderated duration judgment magnitude. However, these variables were probably significant moderators mainly because they were correlated with duration judgment method. These variables included duration length, stimulus modality, participants' sex, and total number of duration judgments.

Interestingly, the magnitude of children's duration judgments is similar to that of old adults' judgments (Block, Zakay, & Hancock, 1998). It is unlikely, however, that the same kinds of duration judgment processes operate in young children as in old adults, because somewhat different variables moderated the effect sizes.

The interindividual variability in duration judgments, as assessed by the coefficient of variation, also showed age-related differences. The mean coefficient of variation decreased greatly from children to adolescents and from adolescents to young adults, regardless of which duration judgment method was used.

Only one experimental report contained data on the psychophysical function relating judged duration to target duration, and from this single study we have drawn no conclusions.

Theoretical Accounts

The present findings reveal substantial age-related differences in the magnitude of verbal estimates and reproductions, along with a consistent age-related reduction in interindividual variability. What kind of theory can handle these findings? Piaget's (1946/1969) account of how children learn to judge durations and why their judgments differ from those of adults only vaguely predicts an age-related difference in duration judgment magnitude. It does not predict (or even postdict) the specific findings that children make larger verbal estimates and shorter reproductions than do adolescents and young adults. However, if different children learn to judge durations in a reasonably accurate way at different rates or ages, it can explain the greater interindividual variability that their judgments show.

Piaget's theory does not offer sufficiently specific predictions concerning processes involved in the development of the ability to judge durations in a relatively accurate way. More recent models, such as the attentional-gate model (Zakay & Block, 1996, 1997), fare better than Piaget's model in that they offer some possible explanations. We now examine how considering some of the more specific physiological and cognitive components may explain the findings.

Potential Physiological Explanations

Pacemaker rate. One way of attempting to explain the differences between children and young adults is to assume that duration judgments are proportional to the rate of a pacemaker or other similar component of an internal timer (see, for example, Zakay & Block, 1996, 1997). Studies measuring reaction time show that children are slower than are young adults (e.g., Nettelbeck & Wilson, 1994), suggesting that their pacemaker rate may be slower than that of young adults. If this is the case and if duration judgments are directly proportional to pacemaker rate, children's verbal estimates should have been smaller, not larger, than those of older participants. Even though it seems to have made the wrong prediction, the pacemaker rate hypothesis may still be viable, however: In order to learn to judge durations accurately, a child may need continually to recalibrate the translation between subjective duration units and verbal estimates. But the recalibration process may lag behind the actual increase in pacemaker rate. For example, if a child experiences a 10-s duration during which 100 subjective temporal units (e.g., pacemaker pulses) occur, he or she may store in reference memory the information ''100 units = 10 s.'' If the pacemaker rate increases throughout childhood, recalibration of reference memory information is needed, but this may lag behind the increase in pacemaker rate. For example, a 10-s duration now may contain 120 subjective temporal units, but the child has not yet revised the reference memory enough. Given no revision, he or she will judge a 10-s duration as being 12 s. For this reason, children may give overly large verbal estimates of durations and to make overly short productions of durations. The present findings, in fact, are in that direction (although the effect size for productions is small and not significantly different from 0). Note that the pacemaker-rate hypothesis is not effectively tested by conditions using the method of reproduction, because any age-related difference in pacemaker rate or calibration during the target duration is also present during the reproduced duration. As such, a simple pacemaker-rate explanation is not able to handle the present finding of an age-related difference in reproduction magnitude.

Temperature or metabolic rate. Brain temperature or basal metabolic rate may affect a central time-generating mechanism, such as a pacemaker. To account for age-related changes in duration judgments, we need to observe corresponding age-related changes in temperature or metabolism. In fact, temperature and metabolism decrease monotonically across the lifespan (Altman & Dittmer, 1968). Further, experienced duration tends to lengthen as body temperature increases (Hancock, 1993). Perhaps children make relatively large verbal estimates because an internal pacemaker is faster, not slower, in children than in adolescents and young adults. However, like children, older adults also make larger verbal estimates than do young adults (Block et al., 1998), so the lifespan trend in duration judgments is not monotonic as the trends in temperature and metabolic rate are. It is also difficult to construct a simple causal account of how temperature or metabolism interacts with duration judgment method, the most important moderator of duration judgment magnitude. Thus, although temperature or metabolism may play some role in duration judgments, they are not the only important influences. At least some sort of hybrid model, containing cognitive components, is needed.

Potential Cognitive Explanations

Duration units. Children may differ from young adults in the translation between subjective and objective duration units. Like the other explanations, this one is post hoc. There is no reason to expect developmental differences in the use of units like seconds and minutes to be in a direction consistent with the present findings. In addition, if children make larger verbal estimates simply because they tend to use larger numbers, such effects would be greater at shorter durations (because of a ''floor effect'' on possible numbers), and we found the reverse. Finally, we are not aware of any evidence that children tend to use larger numbers than do young adults when they are estimating other magnitudes. Some evidence indicates the opposite: Compared to young adults, children make smaller estimates of numerosity of dots in a visual display (Ginsburg, 1994).

Memory processes. Another possibility is that there are developmental differences in memory processes, such as the rate of forgetting of information. In this kind of explanation, the observed differences between participants of different ages may reflect a more rapid loss of information concerning events near the start of a time period. If children forget those events at a greater rate than do young adults, they may infer that the duration is relatively longer. (An event that is remembered less well may seem to have occurred longer ago.) Thus, their verbal estimates should be larger and their productions shorter. In the case of reproductions, forgetting occurs at the same rate during the target duration and during the reproduced duration, so we expect no age-related differences. Because this is not what the reproduction data show, a simple forgetting explanation is not tenable. Age-related differences in memory processes may be more important in the retrospective paradigm (i.e., in studies of remembered duration) than in the prospective paradigm (i.e., in studies of experienced duration); as we have noted, all extant evidence came from use of the prospective paradigm.

Attentional resources. The attentional-gate model of duration judgment emphasizes the influence of attentional resource allocation on prospective duration judgments (Zakay & Block, 1996, 1997). Children may have more limited attentional resources than young adults have. Consequently, many everyday tasks may be so attention-demanding that children ordinarily have few attentional resources with which to attend to time. Young adults, on the other hand, may ordinarily have sufficient resources with which to attend to time. In the experiments reviewed here, typically there was either no task or a very easy concurrent (nontemporal) information-processing task during the time period. Compared to everyday situations, children may have had relatively greater residual resources with which to attend to time. Because verbal estimates and productions must rely on correspondences between objective and subjective units learned in everyday situations, this would tend to lengthen their verbal estimates and shorten their productions. (Effects on reproductions would depend on whether or not the task difficulty differed between the target duration and the reproduced duration.) This explanation predicts that nontemporal task difficulty should be an important moderator variable, along with duration judgment method. Unfortunately, no researcher has investigated this variable in an experiment comparing children and older participants (see, however, Arlin 1986a, 1986b). Nearly all of the experiments we reviewed here contained only a simple information-processing task during the target duration. Future studies should investigate this variable.

Impatience and waiting. Compared to adolescents and young adults, children may have more difficulty waiting for something to happen (Fraisse, 1982). While a target duration is elapsing, children may be relatively impatient, focusing on the time at which the duration will end. This may increase the amount of attention they allocate to time, thereby lengthening their prospective duration judgments (Zakay & Block, 1996, 1997). Children's verbal estimates may have been relatively large for this reason. In addition, when children attempt to reproduce a duration, they may be similarly impatient and thereby terminate the reproduction sooner than adolescents and young adults do. Of all the potential explanations for the present findings, this one seems the most feasible. Thus, an attentional-gate model can at least offer a post hoc explanation in its ability to handle hypothetical mediator variables such as impatience.[9]

CONCLUSIONS

Considering the pattern of results we have reported here, along with findings of other age-related differences (Block et al., 1998), it is unlikely a single hypothesis can account for all developmental differences in duration judgments. Age-related differences in duration judgment ratio and coefficient of variation may be attributable to a number of processes. In particular, the present evidence suggests that: (a) learning to give numerical estimates of durations in a reasonably accurate way continues throughout childhood, and (b) children are impatient or unable to delay a response during relatively empty durations. The finding of greater interindividual variability in children than in adolescents and young adults, which was consistently found for all duration judgment methods, strongly suggests that children develop time-related abilities at different rates or ages. This seems to be especially the case concerning impatience (or the inability to delay a response): In the method of reproduction, a child has to wait before signaling the end of the reproduction, and conditions using that method showed greater developmental differences in interindividual variability than did those using other methods.

[9] The direction of the causal link between impatience and duration judgments is unclear. Children's temporal experience may be inherently distorted, perhaps because they allocate relatively greater attention to time, and impatience may result from this distortion. We are indebted to Teresa McCormack for this suggestion.

The role of physiological variables, such as brain temperature and its possible effect on the rate of an internal pacemaker, has not yet been adequately researched in a developmental framework. The potential role of these variables is unclear. Admittedly, it is difficult to conduct appropriate research with children, because the required physiological manipulations are invasive.

Whether children differ from adolescents and young adults in retrospective duration judgments is also unclear, because no study has used this paradigm. Memory processes tend to be critically important in the retrospective paradigm (Block, 1990). Experiments comparing retrospective duration judgments made by children and adults are critically important. If children tend simply to use larger numbers when they make verbal estimates, retrospective judgments should show the same age-related difference as we found here with prospective judgments. On the other hand, children may make shorter retrospective verbal estimates if they have a relatively poor ability to remember events from the target duration.

Future research should test hypotheses about developmental differences by using much more varied experimental conditions (duration judgment paradigm, amount of stimulus information presented, processing task difficulty, and so on). Such research is needed in order to reach any truly definitive conclusions. More complete developmental research may also serve the important function of clarifying models of duration judgment processes in adults, thereby providing a more complete picture of the results of childhood learning about time.

REFERENCES

References marked with an asterisk (*) indicate studies included in one or both of the meta-analyses.

Altman, P. L., & Dittmer, D. A. (1968). *Metabolism.* Bethesda, MD: Federation of the American Societies for Experimental Biology.

Arlin, M. (1986a). The effects of quantity and depth of processing on children's time perception. *Journal of Experimental Child Psychology,* **42,** 84–98.

Arlin, M. (1986b). The effects of quantity, complexity, and attentional demand on children's time perception. *Perception & Psychophysics,* **40,** 177–182.

Arons, M., & London, H. (1969). Correcting for compensation in studies of time estimation. *Psychonomic Science,* **17,** 319–320.

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

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

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. (1998). *The compleat bibliography on the psychology of time, 1839–1997* [Machine-readable data file]. Bozeman, MT: Department of Psychology, Montana State University. [Producer and distributor].

Block, R. A., & Pierce, C. A. (1998). *Using primary-level statistics to improve meta-analytic interpretations.* Manuscript in preparation.

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

Cahoon, R. L. (1969). Physiological arousal and time estimation. *Perceptual and Motor Skills,* **28,** 259–268.

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

Dmitriyev, A. S. (1980). [Human orientation in time (conscious estimation of short time intervals)]. *Uspekhi Fiziologicheskikh Nauk,* **11,**, 47–67.

*Dmitriyev, A. S., & Tushnova, T. V. (1967). [On the change in the conditioned reflex to time after the school day in children of various school age]. *Zhurnal Vysshei Nervnoi Deiatel'nosti,* **17,** 413–418.

*Dmitriyev, A. S., & Voitiukova, Z. V. (1973). [Formation of so-called subjective estimation of time]. *Zhurnal Vysshei Nervoi Deiatel'nosti,* **23,** 704–710.

Droit-Volet, S. (1998). Time estimation in young children: An initial force rule governing time production. *Journal of Experimental Child Psychology,* **68,** 236–249.

Eisler, A. D., Eisler, H., Guirao, M., & Harris, P. (1995). Time perception in children. In C.-A. Possamaï (Ed.), *Fechner Day 95.* Cassis, France: International Society for Psychophysics.

Elkine, D. (1928). De l'orientation de l'enfant d'age scolaire dans les relations temporelles. *Journal de Psychologie Normal et Pathologique,* **25,** 425–429.

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

*Fraisse, P. (1948). Étude comparée de la perception et de l'estimation de la durée chez les enfants et les adultes [Comparative study of the perception and estimation of duration in children and adults]. *Enfance,* **1,** 199–211.

Fraisse, P. (1982). The adaptation of the child to time. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 113–140). New York: Academic Press.

Fraisse, P., & Orsini, F. (1958). Étude expérimentale des conduites temporelles: III. Étude génétique de l'estimation de las durée [Experimental study of temporal behavior: III. Genetic study of the estimation of duration]. *Année Psychologique,* **58,** 1–6.

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

Friedman, W. J. (1978). Development of time concepts in children. *Advances in Child Development and Behavior,* **12,** *267–298.*

*Gareyev, E. M. (1977). [The specifics of formation of subjective time estimation during ontogenesis]. *Voprosy Psikhologii,* **5,** 114–120.

*Gareyev, E. M., & Osipova, L. G. (1980). [Age and time estimation during different types of activity]. *Zhurnal Vysshei Nervoi Deiatel'nosti,* **30,** 251–255.

*Gilbert, J. A. (1894). Researches on the mental and physical development of school-children. *Studies from the Yale Psychology Laboratory,* **2,** *40–100.*

Gilliland, A. R., & Humphreys, D. W. (1943). Age, sex, method, and interval as variables in time estimation. *Journal of Genetic Psychology,* **63,** 123–130.

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

Ginsburg, N. (1994). Numerosity estimation and age. *Perceptual and Motor Skills,* **79,** 1434.

*Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1958). Kinesthetic cues in the development of time concepts. *Journal of Genetic Psychology,* **93,** 185–190.

Goldstone, S., & Goldfarb, J. L. (1966). The perception of time by children. In A. H. Kidd & J. L. Rivoire (Eds.), *Perceptual development in children* (pp. 445–486). New York: International Universities Press.

Goldstone, S., Lhamon, W. T., & Boardman, W. K. (1957). The time sense: Anchor effects and apparent duration. *Journal of Psychology,* **44,** 145–153.

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

Guyau, J.-M. (1988). La genèse de l'idée de temps [The origin of the idea of time]. In J. A. Michon, V. Pouthas, & J. L. Jackson (Eds.), *Guyau and the idea of time* (pp. 37–90; trans. pp. 93–148). Amsterdam, Netherlands: North-Holland. [Original work published 1890]

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

Hancock, P. A., & Manser, M. P. (1997). Time to contact: More than tau alone. *Ecological Psychology,* **9,** 265–299.

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

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

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

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

Herman, J. F., Norton, L. M., & Roth, S. F. (1983). Children and adults' distance estimations in a large-scale environment: Effects of time and clutter. *Journal of Experimental Child Psychology,* **36,** 453–470.

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

*Hicks, R. E., Allen, D. A., & Mayo, J. P. (1984). A developmental study of temporal duration judgments. *Journal of Genetic Psychology,* **144,** *31–38.*

Hornstein, A. D., & Rotter, G. S. (1969). Research methodology in temporal perception. *Journal of Experimental Psychology,* **79,** 561–564.

Iberall, A. S. (1992). Does intention have a characteristic fast time scale? *Ecological Psychology,* **4,** 39–61.

James, W. (1890). *The principles of psychology.* (Vol. 1). New York: Henry Holt.

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.

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 Psicologia,* **4,** 35–41.

LeBlanc, A. [F.] (1966). Orientation and estimation of time as a function of age. *Dissertation Abstracts International,* **27,** 1624–1625.

*LeBlanc, A. F. (1969). Time orientation and time estimation: A function of age. *Journal of Genetic Psychology,* **115,** 187–194.

*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. (1974). Effects of space and velocity on time estimation in children and adults. *Psychological Research,* **37,** 107–123.

Matsuda, F. (1989). A developmental study on a duration estimation: Effects of frequency of intermittent stimuli. *Japanese Psychological Research,* **31,** 190–198.

*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.*

*Matsuda, F., & Matsuda, M. (1976). Effects of frequency of intermittent stimuli on time estimation in children and in adults: II. *Psychologia: An International Journal of Psychology in the Orient,* **19,** *11–22.*

Matsuda, F., & Matsuda, M. (1987). A developmental study of the role of learning processes in duration estimation. *Japanese Psychological Research,* **29,** 27–36.

Matsuda, F., Miyazaki, H., & Matsuda, M. (1983). Effects of style of response on learning of time estimation in young children and adults. *Bulletin of the Faculty of Home Economics, Hiroshima Women's University,* **19,** 41–48.

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

Nettelbeck, T., & Wilson, C. (1994). Childhood changes in speed of information processing and mental age: A brief report. *British Journal of Developmental Psychology,* **12,** 277–280.

*Orihara, S. (1980). The effect of mental image in time estimation. *Journal of Child Development,* **16,** 13–20.

Piaget, J. (1969). *The child's conception of time* (A. J. Pomerans, Trans.). New York: Basic Books. [Original work published 1946]

Pigott, T. D. (1994). Methods for handling missing data in research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 163–175). New York: Russell Sage Foundation.

Pöppel, E. (1988). *Mindworks: Time and conscious experience* (T. Artin, Trans.). Boston: Harcourt Brace Jovanovich. [Original work published 1985]

Pouthas, V. (1993). Ontogenesis of temporal learning in the child: Experimental evidence and perspectives. *Psychologica Belgica,* **33,** 171–183.

Richards, W. (1964). Time estimates measured by reproduction. *Perceptual and Motor Skills,* **18,** 929–943.

Richards, W. (1973). Time reproductions by H.M. *Acta Psychologica,* **37,** *279–282.*

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

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review,* **17,** 881–901.

Smythe, E. J. (1956). A normative, genetic study of the development of time perception. *Dissertation Abstracts International,* **16,** 2517.

Smythe, E. J., & Goldstone, S. (1957). The time sense: A normative, genetic study of the development of time perception. *Perceptual and Motor Skills,* **7,** 49–59.

Tejmar, J. (1962). Age differences in cyclic motor reaction. *Nature,* **195,** 813–814.

Tracz, S. M. (1985). The effect of the violation of the assumption of independence when combining correlation coefficients in a meta-analysis (Doctoral dissertation, Southern Illinois University, 1984). *Dissertation Abstracts International,* **46,** 688A.

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

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the ''internal clock.'' *Psychological Monographs,* **77** (13, Whole No. 576), 571–513.

Wallon, H., Evart-Chmielniski, E., & Denjean-Raban, G. (1957). Reproduction de durées courtes par l'enfant [Reproduction of short time intervals by the child]. *Enfance,* **2,** 97–134.

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

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

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science,* **6,** 12–16.

Received: June 12, 1998; revised: September 28, 1998