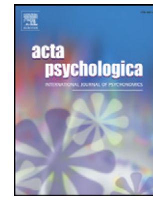
Acta Psychologica 134 (2010) 330–343

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy



How cognitive load affects duration judgments: A meta-analytic review

Richard A. Block [a,*], Peter A. Hancock [b], Dan Zakay [c]

[a] Department of Psychology, Montana State University, Bozeman, MT 59717-3440, United States
[b] Department of Psychology, University of Central Florida, Orlando, FL 32816-1390, United States
[c] New School of Psychology, Interdisciplinary Center Herzliya, and Department of Psychology, Tel Aviv University, Ramat Aviv 69 978, Tel Aviv, Israel

ARTICLE INFO

Article history:
Received 1 July 2009
Received in revised form 18 March 2010
Accepted 19 March 2010
Available online 18 April 2010

PsycINFO classification:
2300

Keywords:
Time estimation
Attention
Memory

ABSTRACT

A meta-analysis of 117 experiments evaluated the effects of cognitive load on duration judgments. Cognitive load refers to information-processing (attentional or working-memory) demands. Six types of cognitive load were analyzed to resolve ongoing controversies and to test current duration judgment theories. Duration judgments depend on whether or not participants are informed in advance that they are needed: prospective paradigm (informed) versus retrospective paradigm (not informed). With higher cognitive load, the prospective duration judgment ratio (subjective duration to objective duration) decreases but the retrospective ratio increases. Thus, the duration judgment ratio differs depending on the paradigm and the specific type of cognitive load. As assessed by the coefficient of variation, relative variability of prospective, but not retrospective, judgments increases with cognitive load. The prospective findings support models emphasizing attentional resources, especially executive control. The retrospective findings support models emphasizing memory changes. Alternative theories do not fit with the meta-analytic findings and are rejected.

1. Introduction

People are experiencing increasing perceptual, attentional, and performance load: automobile drivers experience cognitive load attributable to the use of cell phones and in-vehicle devices and to increased traffic. Airplane pilots and air-traffic controllers experience cognitive load attributable to complicated instrumentation and to increased air traffic. The increasing complexity of technology makes high cognitive load ubiquitous (Hancock & Szalma, 2008). Ways to reduce errors in human performance under conditions of cognitive load depend on methods to measure load. Those measurements are derived from and have an impact on basic theories of human attention, perception, and performance and how they are best assessed. As information processing increasingly became the focus of modern work, there arose a need to identify methods to evaluate cognitive load. To resolve this problem, researchers looked to previous methods of physical work assessment for solutions. One approach, first developed in time-and-motion studies at the turn of the last century (e.g., Taylor, 1913), is primary task performance. This takes the form of online measurement of output in relation to the task that people perform. If a task involves industrial processing, the number of units of a product per unit of time reflects the level of load experienced. Unfortunately, for many tasks, the output rate is difficult to specify.

However, secondary task performance is the one most rooted in psychological theory. Largely founded on the notion of limited attentional capacity, this methodology argues that as the cognitive load demanded by performance of a primary task increases, the performance on a secondary task decreases. Ways to measure cognitive load include physiological measures, primary task performance, secondary task performance, and opinion surveys. Attentional resource theories (Kahneman, 1973; Navon & Gopher, 1979; Wickens & Kessel, 1980) focused on secondary task methodology, using tasks that presumably required the same attentional resources as the primary task. The undifferentiated attentional resource model, first proposed by Kahneman, rendered this assessment process simple because all cognitive tasks were assumed to compete for a single limited pool of attentional resources. However, when subsequent theorists proposed multiple resource pools, the choice of a specific secondary task became problematic. Questions arose as to which secondary tasks tapped which respective resource pools, and evidence began to accumulate of dissociations between increasing task difficulty and primary and secondary task performance (Hancock, 1996). Although cognitive load measures usually agree, instances of dissociation reveal the lack of theoretical guidelines as to when they might occur (but see Yeh & Wickens, 1988).

1.1. Duration judgments as a cognitive load measure

Time (duration) estimation, a measure of secondary task performance, has been shown in several experiments to be a reliable and valid measure of cognitive load. For this reason, applied researchers, beginning with Hart (1975) and Casali and Wierwille (1983, 1984),

* Corresponding author. Tel.: +1 406 994 5173.
E-mail address: block@montana.edu (R.A. Block).

0001-6918/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.03.006

have increasingly focused on duration estimation. It is thought that when a person is working on a difficult or attention-demanding task, time seems to pass quickly, but if a person is working on an easy or less attentional-demanding task, time seems to pass slowly (Block, George, & Reed, 1980; Block & Zakay, 2008; Brown, 2008). Although the past century of research contains findings that support these intuitive observations, researchers have failed to reveal the reasons for these kinds of temporal distortion. Our meta-analytic review focuses on the first century of research on this issue, which dates from the seminal study of Yerkes and Urban (1906). It establishes the relative size and direction of these effects, and it also tests various models that have been proposed to explain the underlying phenomena.

Reasons to investigate the effects of cognitive load on human duration judgments are motivated by both basic and applied concerns. Understanding the effects of cognitive load on duration judgments can help develop and refine theories of human duration judgment and, more generally, human information processing. For example, one current hypothesis is that cognitive load is "a function of the proportion of time during which a given activity captures attention, thus impeding other central processes" (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007, p. 570). Along these lines, several researchers (e.g., Rammsayer & Brandler, 2007; Zakay, Block, & Tsal, 1999) have suggested that duration judgments may be a reliable and valid index of cognitive load to the extent that they involve time-shared central processes, especially attentional, executive, or working-memory resources. Recent research on duration judgment processes has increasingly focused on issues concerning the division of attentional resources between nontemporal and temporal information processing (for recent reviews, see Block, 2003; Grondin, 2001, 2008; Zakay & Block, 1997).

We conducted a meta-analysis focusing broadly on the effects of cognitive load on human duration judgments. A critical feature of it is the specification of the meaning of the term *cognitive load* and how it is used in the literature. We define *cognitive load* as the amount of information-processing (especially attentional or working-memory) demands during a specified time period; that is, the amount of mental effort demanded by a primary task. These demands may also include some heavily cognitively driven perceptual-motor processes. We use the term *cognitive load*, or more simply *load* (e.g., Barrouillet et al., 2007) instead of various near-synonymous terms that other researchers use, such as *mental workload* (e.g., Hancock & Meshkati, 1988; Proctor & Van Zandt, 1994; Wierwille, Rahimi, & Casali, 1985), *cognitive workload* (e.g., Patten, Östlund, Joakim, Nilsson, & Svenson, 2006), or simply *workload* (e.g., Gopher & Donchin, 1986).

1.2. Theoretical focus

In most experiments on duration estimation, researchers have obtained duration judgments as a function of nontemporal stimulus information or information-processing demands, not merely as a judgment of an *empty* duration (one devoid of externally presented stimulus content). In one early study, Swift and McGeoch (1925) asked college students to judge the duration of a time period during which they either listened to an interesting passage (a low-load condition) or wrote down the passage while they were listening to it (a high-load condition). Many early researchers (e.g., Gulliksen, 1927) used several qualitatively different kinds of tasks, with the type of activity apparently selected in an atheoretical way. One critical issue is that there has been little if any theoretical coherence in the choice of tasks used to manipulate cognitive load. Our meta-analysis remedies this failure by classifying different kinds of cognitive load and assessing each type separately, then relating the findings to theoretical accounts.

One of the earliest reviews of the literature on time perception and estimation was written by Weber (1933). His review was essentially a narrative summary of wide-ranging articles (49 of them) on the psychology of time. For present purposes, he distinguished between the amount of mental content and its complexity, which others have mentioned more recently (e.g., Ornstein, 1969). Even more recent duration judgment literature contains reports of some experiments in which participants passively viewed different numbers or complexities of stimuli, or in which they only estimated time or estimated time while performing a task, without any manipulation of the load of that task. Although it might be argued that passively viewing fewer or less complex stimuli requires lower load demands, whether participants increase their cognitive load if they passively view more stimuli or more complex stimuli is unclear. In these experiments (e.g., Ornstein, 1969), sensory or perceptual factors could influence any observed differences in duration judgments, not cognitive load per se. Therefore, we have not included those kinds of studies in the present meta-analysis. As much as possible, we included only comparisons of experimental conditions in which the number of presented stimuli were comparable in high and low load conditions, and in which the main manipulation involved cognitive load per se, not merely a single-task (timing only) condition compared to a dual-task (timing plus a secondary task) condition. In some studies that made that comparison, in the dual-task conditions participants were instructed to make verbal responses that created a sensory–perceptual "filling" of the interval. According to the well-known *filled-duration illusion* (see, for example, Poynter, 1989), duration estimates lengthen if a duration is filled, as opposed to unfilled. In order to determine whether cognitive load per se affects duration estimates, we needed to rule out the possibility of an artifactual change in duration judgments attributable to the filled-duration illusion.

1.3. Duration judgment paradigm: models and predictions

Some studies have revealed that the duration judgment paradigm affects duration estimates. In the *prospective* paradigm, a person is aware prior to or immediately upon the onset of a duration that a duration judgment is necessary and important. In contrast, in the *retrospective* paradigm, a person becomes aware only after the duration has passed that a duration judgment is needed. The current duration judgment literature reveals several major theoretical controversies: (a) Do the processes underlying prospective and retrospective duration judgments differ and, if so, in what ways? (b) Are prospective duration judgments affected by attentional processes, or can they be explained by other kinds of processes? (c) Are retrospective duration judgments affected by memory processes, or can they be explained by other kinds of processes?

Prospective and retrospective paradigms sometimes show opposite effects on duration judgments (e.g., Block, 1992; Block & Zakay, 1997), but this is by no means an invariable finding. For example, Brown and Stubbs (1992) concluded that "similar timing processes operate under prospective and retrospective conditions" (p. 545). One way to resolve this issue is to determine whether various kinds of cognitive load differentially affect duration judgments under the two paradigms. If cognitive load affects duration judgments in different ways, this finding will finally resolve important theoretical arguments about whether prospective and retrospective duration judgment processes are similar or different.

We first investigated duration judgment paradigm as a potential moderator variable. Because it was, we conducted separate analyses to investigate the specific cognitive load and other moderator variables that affect judgments in each paradigm. Consider the broad context of models of prospective and retrospective time estimation and how they do or do not make predictions about effects of cognitive load.

1.3.1. Prospective paradigm

Treisman (1963) proposed one of the first formal models of an internal clock, which included a pacemaker, a counter, and a comparator mechanism. At present, the most influential pacemaker-accumulator model of prospective timing is scalar expectancy theory (SET). It was proposed to explain the timing behavior of animals such as pigeons and rats, and it remains influential (e.g., Church, 2006; Gibbon, Church, & Meck, 1984). According to SET, animal timing relies on an internal clock that consists of a pacemaker which produces pulses at a regular rate, a switch that controls the passage of the pulses, and an accumulator that counts the pulses. The switch starts and stops the timing of an interval, closing and opening with the perception of stimuli that signal the start or the end of an interval. The properties of the internal clock, according to SET, ensure that the internal clock measures time along a scale that is a monotonic function of clock time.

A major difference between animals' and humans' timing behavior is the role of attention in timing. Whereas this role is minimal in animals (Lejeune, Macar, & Zakay, 1999), it is critical in humans (Brown, 2008). Thus, other researchers proposed the attentional-gate model (AGM; Block & Zakay, 1996; Zakay & Block, 1995), which is an elaboration of the original SET and other similar models. (Subsequently, some theorists have elaborated on SET to propose an attentional influence on the switch.) The AGM includes a critical component, a gate that is opened as a result of attention to time during a duration, not simply a switch that reacts to the start signal and end signal of a duration (or that performs two seemingly different functions, reacting to such signals and being sensitive to cognitive load during a duration). The AGM thus clearly predicts effects of cognitive load. Prospective timing is assumed to demand the same attentional or working-memory resources that nontemporal processing (cognitive load) requires. This creates dual-task interference, which affects duration estimates: as nontemporal processing demands increase, subjectively experienced duration decreases.

Recently, Dutke (2005) argued, based on his and earlier research (e.g., Brown, 1997), that a model such as the AGM, which proposes "a unitary concept of task difficulty and attentional demands" (p. 1411), should be modified. Instead, a more specific resource approach, probably involving the central executive component of working-memory models (e.g., Baddeley, 1986), is needed. The present meta-analysis investigates and clarifies this issue by revealing what specific types of cognitive load influence prospective duration judgments.

Another model, which has been investigated in humans, proposes that prospective timing involves the formation on temporal expectancies based on stimulus rhythm and expected duration endings (e.g., Boltz, 1995, 2005; Jones, Moynihan, MacKenzie, & Puente, 2002). Boltz (1991) provided examples for such patterns in which interactions between temporal expectations and actual endings of target intervals create conditions of early or late endings. Healy, Woldmann, Parker, and Bourne (2005) proposed that prospective timing is not a dual-task condition because the timing and nontemporal tasks can be combined and performed as one task according to what they called a *functional task principle*. None of these kinds of models offers any obvious, explicitly stated way to attribute prospective timing to cognitive load.

1.3.2. Retrospective paradigm

In the retrospective paradigm, people are assumed to not be allocating attentional resources in order to attend to time during a duration while it is in progress. Some influential theories of retrospective timing focus mainly on the nature of the stimulus events during the duration, such as their numerosity, complexity, or memorability (e.g., Ornstein, 1969; see Block (1989, 2003) for reviews). In general, these stimulus-based kinds of theories do not predict that cognitive load per se will affect retrospective duration judgments. More recent models of retrospective duration judgment predict effects of at least some types of cognitive load, although in the opposite direction from those found in the prospective paradigm. For example, the contextual-change model (e.g., Block, 2003; Block & Reed, 1978), as well as a variant of it, the segmentation model (Poynter, 1989) predicts that some types of cognitive load—specifically, increased processing changes or interval segmentation—lengthens remembered duration. According to these models, remembered duration lengthens as the amount of contextual changes or interval segmentation (such as by the inclusion of high-priority events) increases. Thus, our meta-analysis might reveal an effect of some types of cognitive load on retrospective duration judgments.

1.4. Types of cognitive load

Different types of cognitive load manipulations provide a way to distinguish among various experimental conditions (the independent variables used in each experiment). Several experimenters (e.g., Brown, 1985) used different kinds of experimental conditions (e.g., a perceptual-motor task at different levels of difficulty, and also a divided vs. selective attention task), and in some of those cases we classified various pairwise comparisons of conditions in more than one cognitive load category. We classified experimental manipulations of cognitive load according to six categories. Although our classification is grounded in current theories (see later), it is primarily based on distinctions between kinds of independent variables manipulated.

1.4.1. Attentional demands

Our operational definition of *attentional demands* is that participants are simultaneously presented more than one source of nontemporal information and must either divide attention between more than one source (high load) or selectively attend to only one source (low load). For example, Brown (1985, Experiment 2) dichotically presented two word lists, and participants were instructed either to attend to both lists (high load) or to only one list while ignoring the other (low load). If attending to time involves some of the same resources as attending to nontemporal information demands, then prospective duration judgments should be affected by this manipulation, although retrospective judgments, which are based more on memory encoding and retrieval processes, might not be affected.

1.4.2. Response demands

Another load manipulation involves requiring participants either to actively respond to presented information (high load) or simply to passively view or hear identical or comparable information during the duration (low load). Participants in high-load conditions engaged in activities such as writing text from dictation (e.g., Spencer, 1921; Swift & McGeoch, 1925; Yerkes & Urban, 1906) or pressing a response key to classify a stimulus (e.g., Predebon, 1996a,b), whereas participants in low-load conditions simply listened to or viewed the stimuli. The active-responding (high-load) condition entails both sensory–perceptual processes and response-selection and response-execution processes, whereas the passive-perceiving (low-load) condition entails sensory–perceptual processes but not response-selection and response-execution processes. We included any experiment that varied whether or not a person was required to respond actively to presented information. We included only comparisons in which the presented stimuli were comparable in passive viewing and active responding conditions. If there was more than one active responding condition, we used only the easier (i.e., minimally demanding) task for the comparison with a passive viewing condition. We expected that the cognitive load imposed by response-selection and response-execution processes would require attentional resources and perhaps affect prospective duration judgments, but that these processes might not affect retrospective duration judgments.

1.4.3. Familiarity

Increased familiarity with an information-processing task might decrease load as a result of learning of some or all of the component processes (Schneider & Shiffrin, 1977). Familiarity decreases retrieval latency and increases the number of associations that are retrieved with the familiar stimulus. Yonelinas, Otten, Shaw, and Rugg (2005) found that recollection and familiarity rely on different networks of brain regions. In the duration judgment literature, familiarity has been manipulated in two ways. One relies on pre-experimental familiarity. For example, participants in Kowal's (1987, Experiment 1) study listened to familiar musical melodies either played in the normal forward direction (low load) or in an unfamiliar reverse direction (high load). Another relies on familiarity induced during the experiment.

For example, participants in some experiments processed ambiguous paragraphs either without any context-clarifying caption or after being provided a title of the paragraph (e.g., Mulligan & Schiffman, 1979; Predebon, 1984). For us to include an experiment in the *familiarity* category, the familiarity could not simply be a result of low-familiarity on the first of two (or more) trials and high-familiarity on the second of two (or more) trials. The reason is that any effect on duration judgment may have resulted from a time-order effect (Block, 1985) or from the classic lengthening effect (Brown, 1997; Hancock, 1993), not an effect of familiarity per se. For some of the theoretical reasons we noted earlier, we expected that both prospective and retrospective duration judgments might be affected by this manipulation, although perhaps in opposite directions.

1.4.4. Memory demands

Another load manipulation involves instructing some participants to try to remember presented information for a later test (high load, usually called an *intentional-memory* condition), whereas other participants were presented identical or comparable stimulus information during the duration but were not instructed also to remember the information (low load, usually called an *incidental-memory* condition). We expected that this workload type might not affect prospective duration judgments much (but see Fortin & Rousseau, 1998), although that it would affect the memory processes that allegedly underlie retrospective duration judgments.

1.4.5. Processing changes

If a participant was required to change the type of information processing performed during a duration, we assumed that the load was greater than in a condition in which the processing type was constant during the duration. This is consistent with many studies that reveal task switching costs (e.g., Gopher, Armony, & Greenshpan, 2000). For example, in some experiments (e.g., Block & Reed, 1978, Experiment 2; Bueno Martinez, 1990, 1992), some participants alternated between structural and semantic levels of processing (high load), whereas other participants performed either only structural or only semantic processing during the duration (low load). We expected that this workload type might affect prospective duration judgments (because of the attentional demands involved in task switching), as well as retrospective duration judgments (because of increased contextual changes), although in opposite directions.

1.4.6. Processing difficulty

The largest number of experiments included in the present meta-analysis manipulated the level of processing difficulty. For some experiments, the classification was clear because the experimenter or experimenters explicitly used the term *difficulty*. Here are several examples: (a) some researchers (e.g., Buchwald & Blatt, 1974, Exp. 2; Smith, 1969) asked participants to solve difficult anagrams or analogies (high load) or to solve easy anagrams or analogies (low load). (b) Some researchers (e.g., Hicks, Miller, & Kinsbourne, 1976) asked participants to sort playing cards according to a 2-bit (high load), 1-bit, or 0-bit (low load) rule. (c) Some researchers (e.g., Block & Reed, 1978, Experiment 1; Hanley & Morris, 1982) asked participants to classify words according to a semantic rule (high load) or a structural rule (low load); this is a rather common levels-of-processing task in the memory literature (Craik & Lockhart, 1972). For other experiments, the processing difficulty classification was decided by us because the researchers did not use that term. We expected that this workload type might affect prospective duration judgments because of the attentional demands involved in more difficult information-processing tasks, as well as retrospective duration judgments, although perhaps in opposite directions.

1.5. Relative variability of duration judgments

Several researchers have noted that cognitive load effects are not only found in the mean of the duration judgments, but also in the intersubject variability of the same duration judgments (e.g., Brown, 1997, 2006; Wierwille & Connor, 1983). One possible source of differences in variability of duration judgments involves individual differences in working-memory capacity, which is related to the concept of attentional control (Feldman-Barrett, Tugade, & Engle, 2004). Perhaps participants who are relatively low in working-memory capacity or attentional control are relatively more affected by high-load conditions. These ancillary data (see Section 3.5) on the variability of duration judgments are important for basic and applied research.

1.6. Recent reviews

Although the topic has increased in interest, only one review that concerns possible effects of cognitive load (defined as *task demands*) on duration judgments has been reported; Brown (1997) reviewed 80 experiments that focused on "effects of task demands on time judgment performance" (p. 1119). In particular, he investigated what he called the *interference effect*, in which increased task demands disrupt timing, causing perceived time to shorten. He used a vote-counting procedure (Bushman, 1994) that takes into account the significance and direction of any reported effect, but which does not account for either the effect size or the sample size. Brown reported that "only 9 studies (11%) were found that either failed to produce the interference effect or produced an opposite pattern" (p. 1119). Although this approach provides valuable insights, it does not yield quantitative outcomes. One nuance of this approach is that tallies can be made even if effect sizes cannot be calculated; presumably, Brown's analysis included experiments that we could not include in our meta-analysis. He did not categorize the various kinds of task demands, and he did not consider experiments that used a retrospective duration judgment paradigm. Our main objective is to provide a quantitative account of effects of cognitive load on duration judgments in which different cognitive load types are distinguished and analyzed separately.

Grondin (2001) recently reviewed some literature on psychological time, focusing mainly on psychophysical issues (e.g., duration discrimination, Weber's Law) that pertain to the comparison of very short durations (those measured in milliseconds). Although his review provided valuable insights on the time perception literature, it did not emphasize issues concerning cognitive load or concerning perception and estimation of longer durations.

2. Method

2.1. Sample of studies

We searched more than 12,000 references on the psychology of time, including references from two major databases, PsycINFO (1887–2008), using their keywords *time perception* and *time estimation*, and Medline (1966–2008), using their keyword *time perception*; published bibliographies on time research; book chapters and books; and our individual files. We searched for articles that contained such terms as *duration judgment*, *work load*, *human channel capacity*, *mental load*, *cognitive load*, *attention*, *familiarity*, *information processing*, *intentional learning*, *difficulty,* as well as many other similar terms. In addition, we searched Web of Science (i.e., the electronic version of *Science Citation Index* and *Social Sciences Citation Index*) for articles that cited some of the more relevant articles (e.g., Brown, 1997; Zakay & Block, 1997). Finally, we checked the reference lists of relevant articles to determine whether any other studies should be included. We did not include any experiment reported in an unpublished dissertation, conference paper, or technical report, mainly because their selective availability might bias the meta-analyses (e.g., older ones not being retrievable).

Our inclusion and exclusion criteria were exactly the same as those used in our previous meta-analyses of duration judgment literature, which were concerned with issues other than cognitive load—duration judgment paradigm (Block & Zakay, 1997), human aging (Block, Zakay, & Hancock, 1998), and developmental changes (Block, Zakay, & Hancock, 1999). We included only experimental data for which we could accurately estimate an effect size from the reported statistics. We excluded any experiment using participants showing gross psychopathology, an altered state of consciousness (e.g., hypnosis; Kurtz & Strube, 2003), or an unusual physiological condition.

Every included experiment involved normal human participants judging durations predominantly equal to or greater than 3 s, with at least one of the independent variables involving cognitive load as we defined it earlier. If an article reported data on duration judgments that included those less than 3 s and those greater than 3 s, we included only data from the durations that were 3 s or more. Perception and estimation of durations less than about 3 s involves very different processes than of longer durations (for reviews and evidence, see Hancock, Arthur, Chrysler, & Lee, 1994; Pöppel, 1985/1988; Wittmann, 1999). For example, the Weber fraction increases at about 3 s (Getty, 1975; Grondin, 2001), reflecting the end of the psychological present (Fraisse, 1984). Finally, the relatively few duration-estimation studies that used durations less than 3 s necessarily (because of reaction-time issues) used methods such as duration discrimination, which is not our current focus. Thus, our meta-analysis is limited in that it only focused on duration estimates of 3 s and longer.

We also excluded any experiment in which the author or authors: (a) were not clear to us (or to them) whether cognitive load was manipulated, and, if so, in what direction (e.g., Gray, 1982; Postman, 1944); (b) reported only a measure of temporal experience other than judgment of duration (e.g., Watt, 1991); (c) did not report sufficient statistics to estimate an effect size from normal participants (e.g., Kurtz & Strube, 2003; Wierwille & Connor, 1983); (d) reported only absolute-error or another similar accuracy score (e.g., Venneri, Pestell, & Nichelli, 2003) because those data cannot be legitimately analyzed along with duration judgment data; (e) used a sample of less than four participants (e.g., Stern, 1904); (f) did not counterbalance the order of prospective and retrospective experimental conditions and therefore entailed a time-order effect in judgment, and for those experiments (e.g., Newman, 1976) we used only the first judgment.

To address one of the currently contentious issues in meta-analytic techniques, we conducted two parallel analyses. In the first analysis, we excluded any experiment for which an effect size had to be approximated from a within-subjects *t* or *F* value, a procedure that was recommended by Dunlap, Cortina, Vaslow, and Burke (1996). In the second analysis, which yielded nearly identical results and which we report here, we included additional studies, using a recommended way to estimate relevant within-subjects effect size (see Section 2.3).

2.2. Coded variables

We coded any available study characteristic (i.e., potential moderator), including but not limited to the following variables from each experiment and from each within-experiment condition. For each experiment, we first coded the type of load variable that was manipulated. If a factorial design was used and more than one load variable was manipulated, we coded and analyzed them separately, combining them in a mean effect size for the initial overall analysis. In addition, we also coded all non-manipulated load variables: (a) attentional demands (unitary—attending to one stimulus/message only, selective—attending to one of two stimuli/messages, divided—simultaneously attending to two or more stimuli/messages, or not applicable); (b) familiarity (low or high, depending on the extent to which participants were exposed to the stimuli or the task either before the experiment or during an earlier phase of the experiment); (c) memory demands (incidental or intentional); (d) processing changes (none, some, or not applicable); (e) response demands (no overt responding to presented stimuli or active responding required); and (f) processing difficulty or level (easy/structural, moderate, difficult/semantic, or not applicable). We coded this information to investigate between-experiment moderator variables, even if a potential load variable was not manipulated as an independent variable in an experiment.

As in previous meta-analyses of the duration judgment literature, we also coded the following non-load variables: (a) publication year, (b) participants' age (children—8.0–12.9 years of age; adolescents—13.0–17.9 years of age; young adults, such as most samples involving college students—18.0–29.9 years of age; old adults—30.0–59.9 years of age; or older adults—at least 60.0 years of age), (c) participants' sex (female, male, both, or unknown), (d) duration length (short—3.0–14.9 s, moderate—15.0–59.9 s, or long—60.0 s or longer), (e) number of stimuli (none, one, several, or many), (f) stimulus modality (visual, auditory, tactile, or other), (g) stimulus complexity (simple, moderate, complex, or not applicable), (h) stimulus segmentation, such as with high-priority events or interruptions during a duration (low, high, or not applicable), (i) duration judgment method (verbal estimation, reproduction, comparison, repeated production, or analogical/absolute), (j) duration judgment immediacy (relatively immediate—delayed only by brief duration judgment instructions, or delayed), and (k) number of trials.

2.3. Effect size analyses

The authors independently estimated effect sizes, resolving disagreements by discussion. Each effect size was calculated as *g*, the difference between the mean duration judgment given by participants in each paradigm divided by the pooled standard deviation (Hedges & Olkin, 1985), using Johnson's (1989, 1993, Version 1.11) DSTAT program and Borenstein, Hedges, Higgins, and Rothstein's (2006, Version 2.2.027) Comprehensive Meta-Analysis (CMA) program. The effect size was defined as positive if the duration judgment ratio was larger under high- rather than low-load conditions and as negative if the duration judgment ratio was smaller under high- rather than low-load conditions. Effect sizes were calculated separately, whenever possible, for different levels of manipulated variables. To provide a single measure for each experiment, we combined all separately calculated effect sizes; although these differences were usually minor, in any case we calculated the mean of them. Each *g* was then converted to *d* by correcting it for bias (Hedges, 1981; Hedges & Olkin, 1985).

If the reported data included means and standard deviations (or standard errors), *d* was calculated from them. If a between-subjects *t* or *F* value was reported, along with means and standard deviations (or standard errors) shown in a figure, the *t* or *F* value was used to calculate *d*. Effect sizes that were calculated from a reported between-subjects *t* or *F* value from a multi-factor design were adjusted according to the recommendations of Morris and DeShon (1997), using DSTAT to reconstruct the ANOVA table.

Whether or not effect sizes that are estimated from a within-subjects *t* or *F* value (with no report of the correlation between paired observations) should be included in a meta-analysis is controversial. Although Dunlap et al. (1996) recommended that they be excluded, some software (such as DSTAT and CMA) provides a default calculation that implicitly or explicitly assumes a correlation between paired observations of 0.50. We conducted all meta-analyses both including and excluding such experiments. For the sake of completeness, we report meta-analyses with those within-subjects *d*s included. Our findings concerning design as a potential moderator of prospective duration judgments suggest that this decision did not distort the analyses. Because a within-subjects design cannot be used to investigate retrospective duration judgments, design and reporting issues are not of concern for that paradigm.

In six experiments (Brown, 1985; Bueno Martinez, 1992; Bueno Martínez, 1994; Macar, 1996; McClain, 1983; Zakay, 1989, Study 3), researchers manipulated more than one type of load. For the overall analysis, we calculated separate *d* values for each type of load, and then we averaged them. For the sub-analyses of load type, we used the separate *d*s.

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

The homogeneity of each set of *d*s was tested to determine whether the conditions shared a common effect size. If there was heterogeneity of effect sizes, as indicated by the statistic Q_W, we attempted to account for it with coded or manipulated study attributes. Two coded variables, publication year and number of trials, are continuous. These variables were analyzed using a weighted least-squares regression model (Borenstein et al., 2006; see also Hedges, 1982b; Hedges & Olkin, 1985). All other coded attributes were categorical. If there were at least three effect size estimates in each of at least two classes of the moderator variable, we used a random-effects categorical model (Hedges, 1982a; Hedges & Olkin, 1985), as implemented by CMA. In a few cases, we combined two similar classes of a variable or excluded a given class if there were fewer than three effect size estimates *k* in a given class. These techniques yield a between-class effect, similar to a main effect in an ANOVA. The homogeneity of effect sizes within each class was estimated by Q_{wi}. The weighted mean effect size (d_{i+}) was calculated for each category, with each effect size weighted so as to give greater weight to effect sizes that were more reliably estimated. If Q_B was significant and there were more than two classes of the moderator variable, we used CMA to perform post hoc paired contrasts of weighted mean effect sizes (d_{i+}) in each pair of classes.

We analyzed data using the random-effects model, which provides a more conservative solution than the fixed-effects model. The random-effects model is based on the assumption that the experiments are a random sample from a population of experiments, and we are generalizing our conclusions to this population. However, we needed to use the fixed-effects model to provide a measure of heterogeneity within each class (Borenstein et al., 2006). For each meta-analysis, the order in which we list and describe moderator variables reflects our judgment concerning the relative importance (from the most to the least) of each variable. This judgment was based on several criteria: (a) the size and significance of the between-class (Q_B) effect in the relevant categorical model, (b) the completeness of the categorical model as indicated by each within-class heterogeneity of variance (Q_W), and (c) differences in the primary-level statistics, as elaborated in Section 2.4.

2.4. Primary statistical analyses

For experiments for which sufficient data were reported, we also calculated the ratio of subjective duration to objective duration—hereafter called the *duration judgment ratio*—separately for low- and high-load conditions. This is a standard measure calculated and reported in many of the studies. Analyzing this ratio affords a comparison between duration judgments made in experimental conditions that entailed different levels of load. It reverses the commonly found negative correlation between the production method and the other methods (primarily verbal estimation) because the participant's production is the actual (objective) duration that corresponds to the verbally requested (subjective) duration. The mean ratio of high-to-low-load judgments was also calculated. In these calculations, the primary statistics were weighted by sample size. Accumulating these primary-level statistics across experiments conditions clarifies the meta-analytic statistics. Mixed-model ANOVAs were performed on unweighted primary-level statistics to clarify the moderator analyses.

2.5. Relative variability analyses

We also analyzed the experiments that provided information, such as cell means and standard deviations (or standard errors), to determine the relative variability of duration judgments in low- versus high-load conditions. (Means and standard deviations were used from each cell because no researcher reported them separately for each participant.) If the two kinds of judgment differ in magnitude, one cannot simply compare standard deviations, because they typically increase with increasing mean judgment when a ratio scale of measurement is involved (Newell & Hancock, 1985). We instead used the common psychometric measure, the coefficient of variation (CV), which is the standard deviation divided by the mean judgment. The present meta-analysis reveals new information about intersubject variability in duration judgment under the two levels of load. Gilpin's (1993) program COEFVAR was used to calculate a χ^2 value for the difference between CVs with the Bennett–Shafer–Sullivan likelihood-ratio test. DSTAT was used to convert each χ^2 to *d*. For this sub-meta-analysis, the effect size was defined as positive if the CV was larger under high- rather than low-load conditions and as negative if the CV was smaller under high- rather than low-load conditions. Primary-level statistics on CVs were also accumulated and tested. For three experiments that were not includable in the main meta-analysis, only data on CVs were reported, and we calculated the effect size directly from them.

3. Results

3.1. Characteristics of experiments

A total of 117 experiments, reported in 85 separate publications (80 journal articles, 4 book chapters, and 1 book), met all criteria for inclusion in one or both main analyses. A total of 77 of these publications were written in English, 3 in Russian, and 1 in Spanish (later in English). For each decade beginning in 1906, the number of included experiments was 1, 4, 1, 2, 0, 3, 8, 34, 29, and 35. This increasing trend is especially notable during the past three decades, during which the proportion of experiments included in this meta-analysis (0.84) is slightly larger than the corresponding proportion of all psychology articles (0.79) and the proportion of “time estimation” or “time perception” articles (0.76). For data on the latter two proportions, see Block and Zakay (2001).

3.2. Overall analyses of effect sizes and duration judgment ratios

For the preliminary overall analysis, only one effect size was calculated for each experiment. (For some experiments, separate effect sizes were calculated and used later if more than one load variable was manipulated or if separate effect sizes could be calculated for different levels of a potential moderator variable. For the overall analysis, these separately calculated effect sizes were averaged.) Of the 113 overall effect sizes, 59 were calculated or estimated from means and standard deviations or standard errors (either reported in text or estimated from a figure), 42 from a between-subjects *F* or *t* value, 8 from a within-subjects *F* or *t* value, 3 from a reported nonsignificant effect (for which we assumed that $d = 0$), and 1 from an inexact *p* value (a report of a significant effect, for which we assumed that $p = .05$). The sign of each effect size was defined as positive if the duration judgment ratio (i.e., the ratio of subjective to objective duration) was greater in the high-load condition and as negative if the ratio was greater in the low-load condition. We report all data for the overall analysis, as well as for moderator variable analyses within each duration judgment paradigm.

The overall weighted mean effect sizes were similar with the fixed-effects and the random-effects model. However, the 95% confidence intervals (*CI*s) were larger with a random-effects model; in other words, that model is more conservative (Hedges & Vevea, 1998). Using it, the resulting overall weighted mean effect size ($d_+ = -0.29$, 95% $CI = -0.39$ to -0.19, $p < .0001$) revealed a greater duration judgment ratio for low-load than for high-load conditions.

Although the overall weighted mean effect size is significant, it is small (Cohen, 1977). Table 1 shows the results of random-effects analyses for the most theoretically and empirically important moderator variable, duration judgment paradigm. For the prospective paradigm, the mean effect was medium and negative, suggesting that there was a larger duration judgment ratio in low-load conditions than in high-load conditions. For the retrospective paradigm, the mean effect was small and positive, suggesting that there was a larger duration judgment ratio in high-load conditions than in low-load conditions.

As Borenstein (2005) discussed in a recent review, CMA was used to conduct several measures of potential publication bias (and associated dissemination and retrieval biases). Inspection of funnel plots (i.e., sample size as a function of effect size) reveals that there was little or no publication bias in either duration judgment paradigm. In addition, the failsafe Ns reveal that an unreasonably large number of experiments would have to have been unpublished (or not retrieved by us) in order for the effect sizes to become nonsignificant at $p = .05$ (failsafe $N = 9038$ and 273 for the prospective and retrospective analyses, respectively). Begg and Mazumdar's rank correlation test reveals that Kendall's tau = 0.07 and −0.07 [p(two-tailed) = .58 and .37, respectively], for the prospective and retrospective analyses. These analyses strongly suggest that there was no publication bias in the present data.

Analyses of primary statistics test the effects of load and paradigm on the duration judgment ratio. A total of 82 experiments contributed prospective duration judgment ratios, and 31 experiments contributed retrospective duration judgment ratios. (A total of 9 experiments contributed to both paradigms.) Fig. 1 shows the mean duration judgment ratio in low- and high-load conditions separately for the prospective and retrospective paradigms.

A mixed-model ANOVA of duration judgment ratios revealed no main effect of load or paradigm [$F(1,111) = 2.25$ and 0.21, $d = 0.25$ and 0.09, $p = .14$ and .65, respectively]. However, the interaction between load and paradigm was significant [$F(1,111) = 40.74$, $d = 0.40$, $p < .001$], revealing opposite effects of cognitive load on prospective and retrospective duration judgment ratios.

Because these findings show that the effect of cognitive load on duration judgments differ under the prospective and retrospective paradigms, we conducted all subsequent moderator analyses separately for each temporal paradigm. Thus, we tested for differences between the six load types separately for each of the two temporal paradigms, as well as for other potential moderator variables.

3.3. Prospective temporal paradigm

The overall weighted mean effect size in the prospective paradigm was moderated by load type (see Fig. 2).

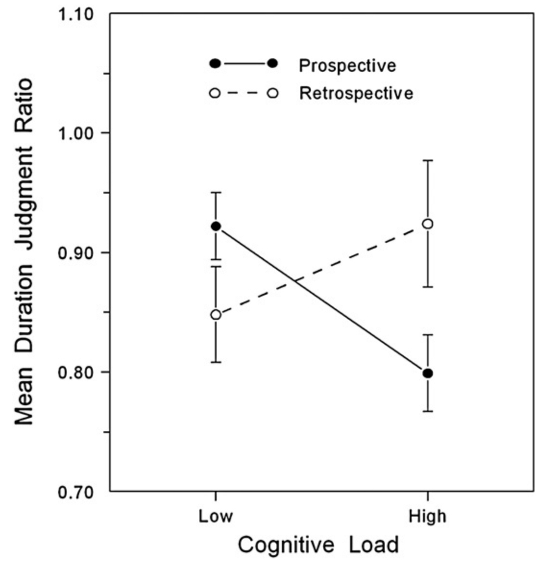


Fig. 1. Mean duration judgment ratio in the prospective and retrospective paradigms as a function of cognitive load. Error bars are the standard errors of the mean.

Three load types showed a significant effect, and these findings are also reflected in the prospective duration judgment ratio:

1. The ratio was smaller if the response demands required active ($M = 0.84$, $SE = 0.05$) instead of passive ($M = 0.99$, $SE = 0.05$) processing [$t(12) = 4.31$, $d = 0.83$, $p = .001$], with an active-to-passive ratio ($M = 0.85$, $SE = 0.04$) of less than 1 [$t(12) = 3.99$, $p = .002$].
2. The ratio was smaller if divided or selective attention ($M = 0.74$, $SE = 0.06$) instead of unitary ($M = 0.84$, $SE = 0.07$) attention was required [$t(7) = 5.09$, $d = 0.59$, $p = .001$], a divided-to-selective/unitary ratio ($M = 0.88$, $SE = 0.02$) of less than 1 [$t(7) = 5.75$, $p = .001$].
3. Processing difficulty. The ratio was smaller if difficult processing ($M = 0.78$, $SE = 0.05$) instead of easy processing ($M = 0.90$, $SE = 0.04$) was required [$t(54) = 6.48$, $d = 0.31$, $p < .001$], with a difficult-to-easy ratio ($M = 0.86$, $SE = 0.02$) of less than 1 [$t(54) = 7.84$, $p < .001$].

Three load types, however, showed no significant effect: memory demands, familiarity, and processing changes. For these three load types, there were no effect size differences between high- and low-load conditions (all $p > .19$), as well as no duration judgment ratio differences (all $p > .23$).

Two other variables, analyzed separately, moderated duration judgments (see Table 2).

Table 1
Overall analyses of duration judgment paradigm as a moderator variable.

Variable and class	Between-classes effect (Q_B)	k	Mean effect size	95% CI Lower/upper	p value	Homogeneity within class (Q_{wi})[a]
Fixed-effects model	243.99***					
Prospective		90	−0.44	−0.48/−0.40	<.001	405.29***
Retrospective		32	+0.28	+0.20/+0.36	<.001	233.97***
Random-effects model	31.83***					
Prospective		90	−0.46	−0.55/−0.37	<.001	
Retrospective		32	+0.26	+0.03/+0.49	.028	

Note. See text for explanations of statistical symbols. Negative effect sizes indicate a greater duration judgment ratio under low- than high-load conditions, and positive effect sizes indicate a greater duration judgment ratio under high- than low-load conditions. Several experiments ($k = 10$) manipulated paradigm between subjects and contributed an effect size in both paradigms.
[a]Significance indicates rejection of the hypothesis of homogeneity.
***$p < .001$.

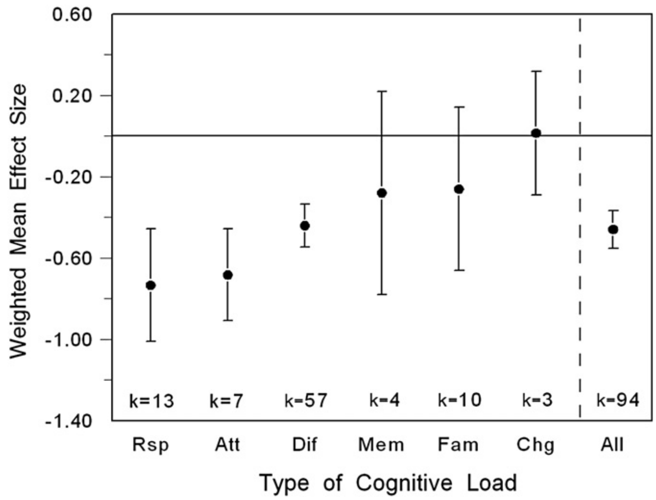


Fig. 2. Weighted mean effect size (and 95% confidence intervals) as a function of cognitive load type in the prospective paradigm: Rsp = response demands, Att = attentional demands, Dif = processing difficulty, Mem = memory demands, Fam = familiarity, Chg = processing changes, All = overall. The letter *k* indicates the number of experiments of each type.

The method by which a particular duration judgment was made moderated the load effect. Although experimental conditions that used the production method showed the largest effect of load, those that used the methods of reproduction and verbal estimation also showed load effects.

Load affected duration judgments if the judgment was made immediately, such as by making a production or by making a verbal estimate or reproduction immediately upon termination of the target duration. However, if participants judged the target duration after a delay of many seconds to several minutes, there was no significant load effect.

Four other variables of particular theoretical interest did not moderate effect sizes.

Experiments using between-subjects and within-subjects manipulation of a load type both showed significant effects ($p<.001$). The effect size was not significantly larger in experiments that manipulated a load type as a between-subjects variable than as a within-subjects variable ($Q_B=2.33$, $p=.13$). In addition, within-subjects experiments for which the effect size was exactly calculated (from means and standard deviations) did not differ in overall effect size from within-subjects experiments for which the effect size was estimated from a reported within-subjects *t* or *F* value. This finding supports our decision to include experiments for which we could only estimate a within-subjects effect size from a *t* and *F* value.

Stimulus duration was not a significant moderator ($Q_B=4.70$, $p=.10$), although weighted mean effect sizes were marginally more negative for short (−0.45) and moderate (−0.56) durations than for long (−0.27) durations.

Stimulus modality was not a significant moderator ($Q_B=1.40$, $p=.24$). The weighted mean effect size was comparable for experimental conditions that used visual stimuli and those that used auditory stimuli. This finding suggests that the effects of cognitive load on prospective duration judgments must be attributed to a central process (e.g., involving attention, working-memory, or both), as opposed to a sensory–perceptual process.

Number of trials was not a significant continuous moderator ($Q_R=0.75$, $p=.38$). We also analyzed number of trials categorically, with experiments that used only one trial (as all retrospective duration judgment conditions necessarily did) compared to those that used more than one trial. In this case, number of trials was a moderator ($Q_B=12.07$, $p=.001$), with experiments using only one trial showing a weighted mean effect size of −0.75 and experiments using more than one trial showing a weighted mean effect size of −0.41. The weighted mean effect size was significant both for both categories ($p<.001$).

Although some findings (e.g., Craik & Hay, 1999) suggest that older adults may be more affected by cognitive load than are younger adults, age was not a significant moderator variable ($Q_B=4.81$, $p=.19$).

3.4. Retrospective temporal paradigm

As we noted earlier, the weighted mean effect size was significantly positive in the retrospective paradigm. Load type did not moderate duration judgments (see Fig. 3).

Table 2
Random-effects analyses of moderators of prospective duration judgments.

Variable and class	Between-classes effect (Q_B)	*k*	Mean effect size	95% *CI* Lower/upper	*p* value	Homogeneity within class (Q_{wi})[a]
Load type	17.66***					
Response demands		13	−0.73	−1.01/−0.46	<.001	75.95***
Attentional demands		7	−0.68	−0.91/−0.46	<.001	6.32
Processing difficulty		57	−0.44	−0.55/−0.34	<.001	225.49***
Familiarity		10	−0.26	−0.65/+0.13	.192	40.91***
Memory demands		4	−0.28	−0.78/+0.22	.271	13.01**
Processing changes		3	+0.01	−0.30/+0.33	.929	3.82
Judgment method	13.79**					
Production		8	−0.75	−1.01/−0.48	<.001	13.51
Reproduction		32	−0.58	−0.74/−0.41	<.001	230.00***
Verbal estimation		38	−0.38	−0.51/−0.26	<.001	150.61***
Repeated production		4	−0.14	−0.40/+0.12	.297	0.76
Judgment immediacy	12.43***					
Immediate		81	−0.51	−0.61/−0.42	<.001	351.254***
Delayed		7	−0.05	−0.25/+0.35	.827	13.33*
Experimental design	0.78					
Between-subjects		45	−0.53	−0.69/−0.37	<.001	199.7***
Within-subjects		44	−0.38	−0.49/−0.28	<.001	158.7***
Exact *d* from *Ms* and *SDs*		30	−0.31	−0.42/−0.20	<.001	91.7***
Estimated *d* from *t* or *F* value		14	−0.57	−0.83/−0.31	<.001	58.6***

Note—See text for explanations of statistical symbols. Negative effect sizes indicate a greater duration judgment ratio under low-load than under high-load conditions.
[a]Significance indicates rejection of the hypothesis of homogeneity.
*$p<.05$.
**$p<.01$.
***$p<.001$.

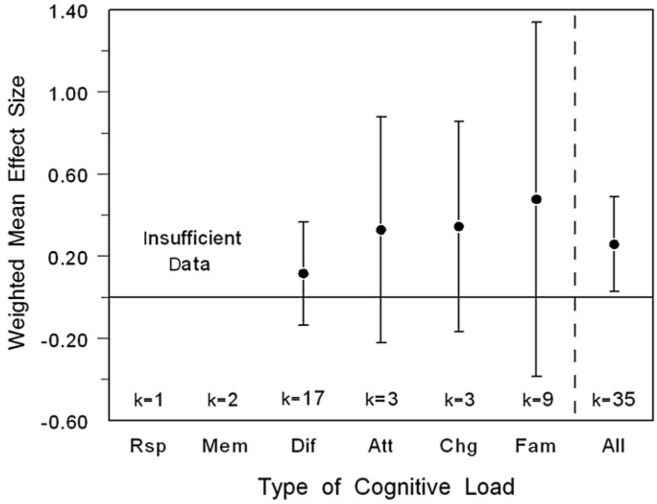


Fig. 3. Weighted mean effect size (and 95% confidence intervals) as a function of cognitive load type in the retrospective paradigm: Rsp = response demands, Mem = memory demands, Dif = processing difficulty, Att = attentional demands, Chg = processing changes, Fam = familiarity, All = overall. The letter *k* indicates the number of experiments of each type.

However, analysis of other coded variables revealed four moderating effects on duration judgments in the retrospective paradigm (see Table 3). Note that these effects should be interpreted as reflecting interaction effects in a (potential or actual) primary experiment.

Experimental conditions with high segmentation during a time period (e.g., high-priority events or interruptions) showed a large, positive load effect on retrospective duration judgments, whereas experimental conditions with little or no segmentation showed no load effect (see, for example, Boltz, 1998). Note that stimulus segmentation differs from processing changes (one of the cognitive load variables) in that no task switching occurred following each high-priority event or interruption.

Conditions that used relatively familiar stimuli during a time period showed a medium-size, positive load effect on retrospective duration judgments, whereas those that used relatively unfamiliar stimuli showed no load effect.

Conditions using a verbal estimation or comparison method showed a medium-size, positive load effect on retrospective duration judgments; whereas those using a reproduction method showed no load effect.

Conditions in which participants judged a long duration (60 s or longer) showed a small-size, positive load effect on retrospective duration judgments, whereas those in which participants judged a short (3.0–14.9 s) or moderate (15.0–15.9 s) duration showed no load effect.

3.5. Coefficient of variation (CV)

The overall weighted mean effect size ($d_+ = +0.12$, 95% $CI = +0.05$ to $+0.18$, $p = .001$) revealed a greater CV for high-load than for low-load conditions. The homogeneity statistic indicated that the overall CV effect sizes were not homogenous [$Q_W(59) = 95.00$, $p = .002$].

Although the overall weighted mean effect size is significant, it is considered small in magnitude (Cohen, 1977). Table 4 shows the results of analyses for the most theoretically and empirically important moderator variable, temporal paradigm.

Paradigm was a significant moderator of CV in the fixed-effects analysis, but it was not a marginally significant moderator in the random-effects analysis. In both analyses, the effect size was significant for experimental conditions that used the prospective paradigm, but it was not significant for experimental conditions that used the retrospective paradigm.

For experimental conditions that used the prospective paradigm, the CV was greater in the high-load ($M = 0.38$, $SE = 0.03$) than in the low-load ($M = 0.34$, $SE = 0.02$) condition [$t(41) = 4.02$, $p < .001$]. In addition, Load Type was a moderator variable [$Q_B(5) = 13.26$, $p = .02$]. The effect size was significant for Response Demands ($k = 6$, $d_+ = +0.38$, 95% $CI = +0.15$ to $+0.62$), and Processing Difficulty ($k = 32$, $d_+ = +0.09$, 95% $CI = +0.02$ to $+0.15$). However, power was low to detect significant effect sizes for the other four Load Types (all $k < 6$, $d_+ < +0.17$, $p > .14$).

For experimental conditions that used the retrospective paradigm, the CV did not differ between the high- ($M = 0.38$, $SE = 0.03$) and low-load ($M = 0.37$, $SE = 0.04$) conditions [$t(20) = 0.43$, $p = .67$]. Load type was not a significant moderator variable [$Q_B(5) = 4.54$, $p = .34$].

4. Discussion

The present meta-analysis was designed to answer several ongoing theoretical controversies about duration judgment processes by investigating possible effects of six types of cognitive load. These controversies concern whether prospective and retrospective duration judgments are based on different processes, and if so, whether attentional (or other) processes can explain prospective duration

Table 3
Random-effects analyses of moderators of retrospective duration judgments.

Variable and class	Between-classes effect (Q_B)	*k*	Mean effect size	95% *CI* Lower/upper	*p* value	Homogeneity within class (Q_{wi})[a]
Segmentation	16.41***					
Little or no		26	+0.03	−0.15/+0.25	.623	92.32***
High		7	+1.29	+0.73/+1.86	<.001	87.15***
Familiarity	5.63*					
Low		7	−0.15	−0.47/+0.18	.371	18.62**
High		16	+0.32	+0.11/+0.54	.003	59.12***
Judgment method	4.97*					
Verbal estimation/comparison		23	+0.42	+0.18/+0.67	.001	146.91***
Reproduction		9	−0.18	−0.65/+0.29	.454	52.96***
Duration length	4.52*					
Short or moderate		8	−0.03	−0.23/+0.17	.761	3.42
Long		24	+0.35	+0.06/+0.63	.018	218.99***

Note—See text for explanations of statistical symbols. Positive effect sizes indicate a greater duration judgment ratio under high-load than under low-load conditions.
[a]Significance indicates rejection of the hypothesis of homogeneity.
*$p < .05$.
**$p < .01$.
***$p < .001$.

Table 4
Analyses of duration judgment paradigm as a moderator of coefficient of variation (CV).

Variable and class	Between-classes effect (Q_B)	*k*	Mean effect size	95% *CI* Lower/upper	*p* value	Homogeneity within class (Q_{wi})[a]
Fixed-effects model	102.74**					
Prospective		45	+0.18	+0.13/+0.23	<.001	73.29**
Retrospective		21	+0.07	−0.05/+0.20	.240	29.45
Random-effects model	0.63†					
Prospective		45	+0.14	+0.07/+0.22	<.001	
Retrospective		21	+0.07	−0.08/+0.22	.351	

Note—See text for explanations of statistical symbols. Positive effect sizes indicate a larger CV under high-load than under low-load conditions.
[a]Significance indicates rejection of the hypothesis of homogeneity.
**$p<.01$.
†$p=.12$.

judgments and whether memory (or other) processes can explain retrospective duration judgments. Our findings reveal how prospective and retrospective duration judgments differ: cognitive load affects duration judgments, but in opposite directions in the prospective and retrospective paradigms. In addition, several cognitive load types affect duration judgments differently in the two paradigms. This provides new information on duration judgments as a measure of cognitive load. Our findings reject theoretical proposals (e.g., Brown & Stubbs, 1992) that similar duration judgment processes operate in the two paradigms and support various theoretical proposals (e.g., Block, 2003; Block & Zakay, 1996) and empirical findings that different duration judgment processes operate in the two paradigms.

When all experiments involving both the prospective and retrospective duration judgment paradigms were analyzed together, the overall weighted mean effect size was significant but small. The overall weighted mean effect size was negative, mainly because researchers have conducted many more studies using the prospective paradigm than using the retrospective paradigm. In the prospective paradigm, the duration judgment ratio (i.e., the ratio of subjective duration to objective duration) decreases under high-load conditions relative to low-load conditions. In the retrospective paradigm, the duration judgment ratio increases under high-load conditions relative to low-load conditions. A second major finding is that several types of cognitive load and other theoretically important moderator variables affected duration judgments differently in the two paradigms. Third, we can rule out a possible artifactual explanation for the finding that duration judgments were affected in opposite directions in the two paradigms; experiments on retrospective duration judgments must necessarily use only one trial, whereas those on prospective judgments may use more than one trial. However, the difference between the two paradigms is seen even in prospective duration judgment experiments that only used one trial.

We first discuss duration judgment processes in the prospective paradigm, then those in the retrospective paradigm.

4.1. Prospective paradigm

The weighted mean effect size in the prospective paradigm was medium and negative. The types of cognitive load manipulations that moderated prospective duration judgments are those that clearly demand working-memory resources, especially those that involve a hypothetical central executive.

If a person had to divide attention between two sources of presented stimuli or selectively attend to one of two sources of presented stimuli instead of merely attending to one source of presented stimuli, the duration judgment ratio decreased. This finding clearly implicates attentional resource allocation models of prospective timing. This is compatible with recent views of attention as a multidimensional cognitive system rather than as a single system (Posner, 2004; Tsal, Shalev, & Mevorach, 2005). This view suggests that there are several attention systems which may function somewhat independently. It is supported by findings concerning attention disorders (e.g., attention-deficit hyperactivity disorder), in which deficits are differently related to various attention systems (e.g., Shalev & Tsal, 2003). Support for this view is also provided by findings showing differential brain correlates of different attention systems (Raz & Buhler, 2006). Divided attention is considered to involve a process subserved by one of these systems. Selective attention is considered to involve another attentional system: executive attention, which enables a person to focus on specific information while ignoring irrelevant or distracting information, like in the case of conjunction search. In that kind of search, participants must respond to two simultaneously present, or not present, features of stimuli (Treisman & Gelade, 1980). Prospective duration judgments are almost always made in an environment containing other concurrent or distracting nontemporal tasks. In order to judge duration accurately, the person needs either to ignore distractors or to divide attention between the timing task and the concurrent task (Parasuraman & Davies, 1984).

If a person had to respond to presented stimuli instead of merely passively perceiving the presented stimuli, the duration judgment ratio decreased. Other evidence, such as that found in studies of the psychological refractory period (Pashler, 1994), suggest that dual-task interference is commonly observed when a person must actively respond to presented stimuli.

If a person had to engage in any of several kinds of relatively difficult processing, as opposed to relatively easy processing, the duration judgment ratio decreased. This finding also supports attentional resource allocation models of prospective timing.

Two other moderator variables also affect prospective duration judgments.

Duration judgment method moderated prospective duration judgments. In studies using the production method, the method that showed the largest load effect, what this means is that under high-load conditions, people make relatively longer productions. This finding supports previous theories that prospective productions are a sensitive, practical, and unobtrusive index of cognitive load (Zakay et al., 1999).

The findings concerning immediacy versus delay of the duration judgment support the argument (e.g., Zakay & Fallach, 1984) that when duration judgment is not immediate, reliance on memory increases and therefore the nature of the judgment process becomes more retrospective than prospective. An alternative explanation is that this finding was driven by the duration judgment method; specifically, the production method, which showed the largest effect of cognitive load, involves immediate instead of delayed judgment.

The finding that similar effect sizes were found in between-subjects and within-subjects designs indicates that the effects of cognitive load on prospective duration judgments are robust. The finding of no effect of stimulus modality implicates central executive processes instead of sensory–perceptual processes.

Three theoretically important cognitive load types apparently do not moderate prospective duration judgments.

Familiarity did not affect prospective duration judgments, although it did affect retrospective duration judgments (see later). Presumably, this is a result of familiarity affecting memory encoding and retrieval, but apparently not attentional resource allocation. Even if processing a familiar stimulus is easier than processing an unfamiliar stimulus, the increased memory search and rate of retrieval associated with familiarity compensates for this by demanding more resources for the enhanced memory processes.

Memory demands were expected to influence retrospective, not prospective, duration judgments because putatively only retrospective duration judgments are affected by memory processes (but see Hancock, 2005). Intentional- as opposed to incidental-memory instructions, lead to a rapid mobilization of attentional resources for long-term memory encoding (Block, 2009). In those five experiments, rapidly presented pictures were recognized better (when they were presented for as few as 0.5 s) under intentional-memory instructions than under incidental-memory instructions.

The lack of an effect of processing changes (i.e., task switching) during a duration on prospective judgments is counter to our original expectation. We expected that processing changes would demand executive control processes and, therefore, attentional resources. It is possible that there was an insufficient number of experiments that manipulated processing changes, or alternatively that processing changes were not strongly varied so as involve sufficient differences in executive control processes. Further research is needed in order to establish whether or not there is an impact of processing changes on prospective duration judgments.

These findings are consistent with attentional models of prospective duration judgment, such as the AGM (Block & Zakay, 1996; Zakay & Block, 1995), and they weaken more simple models on which the AGM model was partly based, such as the original SET model. Under high-load conditions, people have fewer resources to allocate to temporal information processing (i.e., attending to time). As a result, the duration judgment ratio decreases. For example, duration productions lengthen. Central executive processes were revealed in the present findings. This can be explained by future models, such as a modified AGM in which the central executive component of working-memory models (e.g., Baddeley, 1986; Dutke, 2005) is implicated in prospective timing. Thus, the attentionally controlled gate in the original AGM is replaced by a gate that is controlled by central executive processes, resulting in an executive-gate model (EGM). An EGM accounts for the present findings, which pose a serious challenge to models that do not feature a central executive component, such as the SET model and several others (see earlier and Block, 2003).

Some duration judgment researchers working in cognitive load frameworks ranging from theoretical to applied have only reported and analyzed data on CVs (or *SD*s), thereby implying that the variability of prospective duration judgments may be a better measure of cognitive load than are the duration judgments from which they are derived. Our findings do not support this approach: effect sizes were larger for duration judgment ratios than for CVs.

4.2. Retrospective paradigm

Some, although not all, theories of retrospective duration judgment predict that a few cognitive load types might affect retrospective duration judgments, although in the opposite direction from the many effects on prospective duration judgments. We found a small and positive overall effect: the duration judgment ratio increased under high-load conditions relative to low-load conditions. This finding is predicted by theories that focus on the potential impact of cognitive load on memory, such as on the number of contextual changes.

A consideration of variables that moderated the effect of cognitive load on retrospective duration judgments (as well as those that did not) enlightens the possible reasons for the effect.

If the duration had been segmented (such as with high-priority events or interruptions), the duration judgment ratio increased (i.e., the load effect was large and positive). This finding can be explained in terms of a segmentation model (e.g., Poynter, 1989) or in terms of a contextual change model (e.g., Block, 2003; Block & Reed, 1978), which regards segmentation as merely one variable that may lead to contextual changes. Our finding reflects findings in some primary literature that segmentation affects retrospective duration judgments but has little or no effect on prospective duration judgments (Zakay, Tsal, Moses, & Shahar, 1994).

Because familiarity with stimuli facilitates memory retrieval, a positive effect on retrospective duration judgment might be expected. However, familiarity had no effect in the prospective paradigm. Thus, the impact of familiarity may be mediated by memory processes. Our findings reveal that retrospective duration judgments are longer for familiar than for unfamiliar stimuli. Support for the assumption that familiarity increases the encoding and retrieval of information was found in several studies (e.g., Koriat, 1993; Ritter & Heder, 1992). Familiarity also enhances the motivation for memory search (Koriat & Levy-Sadot, 2001; Metcalfe, Schwarts, & Joaquim, 1993).

The moderating effect of duration judgment method can be attributed to the involvement of long-term memory when verbal estimation or comparison methods are used.

The moderating effect of duration length is of interest. If the target duration was less than 60 s, there was no effect of cognitive load in the retrospective paradigm. On the other hand, if the target duration was greater than 60 s, there was a significant effect. Most of the experimental conditions that used short or moderate target durations used one that was less than 20–30 s. This finding suggests that long-term memory retrieval processes are important in the retrospective paradigm. It also suggests that different processes are involved in retrospective judgments of short and long durations.

Although memory type (incidental vs. intentional) could not be assessed because there was only one relevant experiment, memory type does influence recognition memory (Block, 2009). This is worth future investigation. Retrospective duration judgments apparently involve recall (availability) more than recognition (familiarity).

5. Conclusion

Human duration judgments are affected by cognitive load, but they are affected in opposite directions depending on the duration judgment paradigm. If participants are aware that duration judgments must be made (prospective paradigm), a greater cognitive load decreases the subjective-to-objective duration ratio. If participants are not aware that duration judgments must be made until after the duration has ended (retrospective paradigm), a greater cognitive load increases the subjective-to-objective duration ratio. Several theoretically interesting variables moderate these differential effects. Duration judgments, especially those made under the prospective paradigm, seem to be a reliable and unobtrusive way to assess cognitive load.

Acknowledgements

This study was supported in part by grants from the United States–Israel Binational Science Foundation (BSF) and the U. S. Army Multiple University Research Initiative (MURI), "Operator Performance under Stress" (P.A. Hancock, Principal Investigator). We thank Diane Conti, Augusta Duncan, and Bonnie Saxton for their help in literature identification and retrieval. We also thank the action editor (Albert Postma), Scott W. Brown, J.L. Szalma, and three anonymous reviewers for their helpful comments and suggestions on the manuscript.

References[1]

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570–585.

[1] References marked with an asterisk indicate studies included in the meta-analysis.

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

Block, R. A. (1989). Experiencing and remembering time: Affordances, context, and cognition. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 333−363). Amsterdam: North-Holland.

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

Block, R. A. (2003). Psychological timing without a timer: The roles of attention and memory. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 41−59). Göttingen, Germany: Hogrefe & Huber.

Block, R. A. (2009). Intent to remember briefly presented human faces and other pictorial stimuli enhances recognition memory. *Memory & Cognition*, *37*, 667−678.

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

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

Block, R. A., & Zakay, D. (1996). Models of psychological time revisited. In H. Helfrich (Ed.), *Time and mind* (pp. 171−195). Kirkland, WA: Hogrefe & Huber.

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. (2001). Psychological time at the millennium: Some past, present, future, and interdisciplinary issues. In M. P. Soulsby & J. T. Fraser (Eds.), *Time: Perspectives at the millennium (The study of time X)* (pp. 157−173). Westport, CT: Bergin & Garvey.

Block, R. A., & Zakay, D. (2008). Timing and remembering the past, the present, and the future. In S. Grondin (Ed.), *Psychology of Time* (pp. 367−394). Bingley, England: Emerald.

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

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

Boltz, M. G. (1991). Time estimation and attentional perspective. *Perception & Psychophysics*, *49*, 422−433.

Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, *57*, 1080−1096.

*Boltz, M. G. (1998). Task predictability and remembered duration. *Perception & Psychophysics*, *60*, 768−784.

Boltz, M. G. (2005). Duration judgments of naturalistic events in the auditory and visual modalities. *Perception & Psychophysics*, *67*, 1362−1375.

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis—Prevention, assessment and adjustments* (pp. 193−220). New York: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2006). (Version 2.2.027) [Computer software]. Englewood, NJ: Biostat.

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

*Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*, 1118−1140.

*Brown, S. W. (2006). Timing and executive function: Bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition*, *34*, 1464−1471.

Brown, S. W. (2008). Time and attention: Review of the literature. In S. Grondin (Ed.), *Psychology of time* (pp. 111−138). Bingley, England: Emerald.

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

*Buchwald, C., & Blatt, S. J. (1974). Personality and the experience of time. *Journal of Consulting and Clinical Psychology*, *42*, 639−644.

*Bueno Martinez, M. (1990). Efectos de los cambios cognitivos y del esfuerzo de procesamiento sobre los juicios de duración prospectivos y retrospectivos. *Estudios de Psicología*, *45*, 109−127.

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

*Bueno Martínez, B. (1994). The role of cognitive changes in immediate and remote prospective time estimations. *Acta Psychologica*, *85*, 99−121.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 193−213). New York: Russell Sage Foundation.

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, *25*, 623−641.

Casali, J. G., & Wierwille, W. W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, *27*, 1033−1050.

Church, R. M. (2006). Behavioristic, cognitive, biological, and quantitative explanations of timing. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence*. New York: Oxford University Press.

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

Craik, F. I. M., & Hay, J. F. (1999). Aging and judgments of duration: Effects of task complexity and method of estimation. *Perception & Psychophysics*, *61*, 549−560.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671−684.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170−177.

*Dutke, S. (2005). Remembered duration: Working memory and the reproduction of intervals. *Perception & Psychophysics*, *67*, 1404−1413.

Feldman-Barrett, L., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553−573.

Fortin, C., & Rousseau, R. (1998). Interference from short-term memory processing on encoding and reproducing brief durations. *Psychological Research*, *61*, 269−276.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, *35*, 1−36.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics*, *18*, 1−8.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. G. Allan (Eds.), *Annals of the New York Academy of SciencesTiming and time perception*, *Vol. 423*. (pp. 52−77) New York: New York Academy of Sciences.

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

Gopher, D., Armony, L., & Greenshpan, Y. (2000). Switching tasks and attention policies. *Journal of Experimental Psychology: General*, *129*, 308−339.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performanceCognitive processes and performance*, *Vol. 2.* (pp. 1−49) New York: Wiley.

Gray, C. (1982). Duration differences: Attentional demand or time error? *Perception*, *11*, 97−102.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin*, *127*, 22−44.

Grondin, S. (Ed.). (2008). *Psychology of time*. Bingley, England: Emerald.

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

Hancock, P. A. (1993). Body temperature influence on time perception. *Journal of General Psychology*, *120*, 197−215.

Hancock, P. A. (1996). Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload. *Ergonomics*, *39*, 1146−1162.

Hancock, P. A. (2005). Time and the privileged observer. *Kronoscope*, *5*, 177−191.

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

Hancock, P. A., & Meshkati, N. (Eds.). (1988). *Human mental workload*. Amsterdam: North-Holland.

Hancock, P. A., & Szalma, J. L. (Eds.). (2008). *Performance under stress*. Aldershot, England: Ashgate.

*Hanley, J. R., & Morris, N. (1982). Time estimation as a function of recall: A test of Ornstein's theory of temporal judgement. *Current Psychological Research*, *2*, 45−53.

Hart, S. G. (1975). Time estimation as a secondary task to measure workload. *Proceedings of the 11th Annual Conference on Manual Control* (pp. 64−77). Washington, DC: U.S. Government Printing Office.

Healy, A. F., Woldmann, E. L., Parker, J. T., & Bourne, L., Jr. (2005). Skill training, retention and transfer: The effects of a concurrent secondary task. *Memory & Cognition*, *33*, 1457−1471.

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.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486−504.

*Hicks, R. E., Miller, G. W., & Kinsbourne, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology*, *89*, 719−730.

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

Johnson, B. T. (1993). Upgrade documentation and computer program. Hillsdale, NJ: Erlbaum.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*, 313−319.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609−639.

Koriat, A., & Levy-Sadot, R. (2001). The combined contribution of the cue-familiarity and accessibility heuristics to feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 34−53.

*Kowal, K. H. (1987). Apparent duration and numerosity as a function of melodic familiarity. *Perception & Psychophysics*, *42*, 122−131.

Kurtz, R. M., & Strube, M. J. (2003). Hypnosis, attention and time cognition. *International Journal of Clinical and Experimental Hypnosis*, *51*, 400−413.

Lejeune, H., Macar, F., & Zakay, D. (1999). Attention and timing: Dual-task performance in pigeons. *Behavioural Processes*, *45*, 141−157.

*Macar, F. (1996). Temporal judgments on intervals containing stimuli of varying quantity, complexity, and periodicity. *Acta Psychologica*, *92*, 297−308.

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

Metcalfe, T., Schwarts, B., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 851—861.

Morris, S. B., & DeShon, R. P. (1997). Correcting effect sizes computed from a factorial analysis of variance for use in meta-analysis. *Psychological Methods*, *2*, 192—199.

*Mulligan, R. M., & Schiffman, H. R. (1979). Temporal experience as a function of organization in memory. *Bulletin of the Psychonomic Society*, *14*, 417—420.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*, 214—255.

Newell, K. M., & Hancock, P. A. (1985). Forgotten moments: A note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions. *Journal of Motor Behavior*, *16*, 320—335.

Newman, M. A. (1976). Movement tempo and the experience of time. *Nursing Research*, *25*, 273—279.

*Ornstein, R. E. (1969). *On the experience of time.* Harmondsworth, England: Penguin.

Parasuraman, R., & Davies, D. R. (Eds.). (1984). *Varieties of attention.* London: Academic Press.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 230—244.

Patten, C. J. D., Östlund, A., Joakim, N., Nilsson, L., & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis & Prevention*, *38*, 887—894.

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

Posner, M. I. (2004). *Cognitive neuroscience of attention.* New York: Guilford.

Postman, L. (1944). Estimates of time during a series of tasks. *American Journal of Psychology*, *57*, 421—424.

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

*Predebon, J. (1984). Organization of stimulus events and remembered apparent duration. *Australian Journal of Psychology*, *36*, 161—169.

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

*Predebon, J. (1996b). The relationship between the number of presented stimuli and prospective duration estimates: The effect of concurrent task activity. *Psychonomic Bulletin & Review*, *3*, 376—379.

Proctor, R. W., & Van Zandt, T. (1994). *Human factors in simple and complex systems.* Boston: Allyn & Bacon.

Rammsayer, T. H., & Brandler, S. (2007). Performance on temporal information processing as an index of general intelligence. *Intelligence*, *35*, 123—150.

Raz, A., & Buhler, J. (2006). Typologies of attentional networks. *Nature Reviews Neuroscience.*, *7*, 367—379.

Ritter, L. M., & Heder, F. E. (1992). What determines initial feeling of knowing? Familiarity with questions terms, not with the answers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435—451.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1—66.

Shalev, L., & Tsal, Y. (2003). The wide attentional window: A major deficit of children with attention difficulties. *Journal of Learning Disabilities*, *36*, 517—527.

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

*Spencer, L. T. (1921). An experiment in time estimation using different interpolations. *American Journal of Psychology*, *32*, 557—562.

Stern, W. (1904). On estimation, particularly time and space estimation. *Beiträge zur Psychologie der Aussage*, *2*, 32—72.

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

Taylor, F. W. (1913). *The principles of scientific management.* New York: Harper and Bros.

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, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97—136.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs: General and Applied*, *77*, 1—31.

Tsal, Y., Shalev, L., & Mevorach, C. (2005). The diversity of attention deficits in ADHD: The prevalence of four cognitive factors in ADHD versus controls. *Journal of Learning Disabilities*, *38*, 142—157.

Venneri, A., Pestell, S., & Nichelli, P. (2003). A preliminary study of the cognitive mechanisms supporting time estimation. *Perceptual and Motor Skills*, *96*, 1093—1106.

Watt, J. D. (1991). Effect of boredom proneness on time perception. *Psychological Reports*, *69*, 323—327.

Weber, A. O. (1933). Estimation of time. *Psychological Bulletin*, *30*, 233—252.

Wickens, C. D., & Kessel, C. (1980). Processing resource demands of failure detection in dynamic systems. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 564—577.

Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors*, *25*, 1—16.

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors*, *27*, 489—502.

Wittmann, M. (1999). Time perception and temporal processing levels of the brain. *Chronobiology International*, *16*, 17—32.

Yeh, Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, *30*, 111—120.

Yonelinas, A. P., Otten, L., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition Memory. *Journal of Neuroscience*, *25*, 3002—3008.

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

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

Zakay, D., & Block, R. A. (1995). An attentional-gate model of prospective time estimation. In M. Richelle, V. D. Keyser, G. d'Ydewalle, & A. Vandierendonck (Eds.), *Time and the dynamic control of behavior* (pp. 167—178). Liège, Belgium: Universite de Liege.

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

Zakay, D., Block, R. A., & Tsal, Y. (1999). Prospective duration estimation and performance. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 557—580). Cambridge, MA: MIT Press.

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

Zakay, D., Tsal, Y., Moses, M., & Shahar, I. (1994). The role of segmentation in prospective and retrospective time estimation processes. *Memory & Cognition*, *22*, 344—351.

Further Reading

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

*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.

*Arlin, M. (1989). The effects of physical work, mental work, and quantity on children's time perception. *Perception & Psychophysics*, *45*, 209—214.

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

*Boltz, M. G., Kupperman, C., & Dunne, J. (1998). The role of learning in remembered duration. *Memory & Cognition*, *26*, 903—921.

*Broadhurst, A. (1969). Time estimation related to personality, cognitive speed and schizophrenia. *Life Sciences*, *8*, 69—78.

*Brown, S. W. (1998). Automaticity versus timesharing in timing and tracking dual-task performance. *Psychological Research/Psychologische Forschung*, *61*, 71—81.

*Brown, S. W., & Bennett, E. D. (2002). The role of practice and automaticity in temporal and nontemporal dual-task performance. *Psychological Research*, *66*, 80—89.

*Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 600—615.

*Burnside, W. (1971). Judgment of short time intervals while performing mathematical tasks. *Perception & Psychophysics*, *9*, 404—406.

*Chattopadhyay, P. K., Dasgupta, S., & Bhattacharyya, A. K. (1982). Estimation of time in children: Effects of stimulus length, task, and personality. *Child Psychiatry Quarterly*, *15*, 104—108.

*DeWolfe, R. K. S., & Duncan, C. P. (1959). Time estimation as a function of level of behavior of successive tasks. *Journal of Experimental Psychology*, *58*, 153—158.

*Fink, A., & Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence: And time estimation as an index of cognitive load. *Personality & Individual Differences*, *30*, 1009—1021.

*Franssen, V., Vandierendonck, A., & Van Hiel, A. (2006). Duration estimation and the phonological loop: Articulatory suppression and irrelevant sounds. *Psychological Research/Psychologische Forschung*, *70*, 304—316.

*Frauenfelder, K. J. (1980). Creativity and the experience of duration. *Journal of Psychology*, *106*, 27—35.

*Friel, C. M., & Lhamon, W. T. (1965). Gestalt study of time estimation. *Perceptual and Motor Skills*, *21*, 603—606.

*Funke, J., & Grube-Unglaub, S. (1991). Verarbeitungstiefe als Einflußfaktor auf das subjektive Zeitempfinden [Depth of processing as a determinant of subjective time estimation]. *Schweizerische Zeitschrift für Psychologie*, *50*, 129—138.

*Gare[y]ev, E. M. (1973). Study of the estimate of time in different kinds of human activity. *Zhurnal Vysshei Nervoi Deiatel'nosti*, *23*, 1077—1079.

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

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

*Gruber, R. P., & Block, R. A. (2005). Effect of caffeine on prospective duration judgments of various intervals depends on task difficulty. *Human Psychopharmacology*, *20*, 275—285.

*Harton, J. J. (1938). The influence of the difficulty of activity on the estimation of time: II. *Journal of Experimental Psychology*, *23*, 428—433.

*Harton, J. J. (1942). Time estimation in relation to goal organization and difficulty of tasks. *Journal of General Psychology*, *27*, 63—69.

*Hawkins, M. F., & Tedford, W. H., Jr. (1976). Effects of interest and relatedness on estimated duration of verbal material. *Bulletin of the Psychonomic Society*, *8*, 301—302.

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

*Hicks, R. E., Miller, G. W., Gaes, G., & Bierman, K. (1977). Concurrent processing demands and the experience of time-in-passing. *American Journal of Psychology*, *90*, 431—446.

*Hirano, S. (1981). Effect of pause between discrete events on time estimation. *Journal of Child Development*, *17*, 35—39.

*Juhnke, R., & Scott, J. N. (1988). Psychology of computer use: V. Computer use and the experience of time. *Perceptual and Motor Skills*, *67*, 863—870.

*Kowal, K. H. (1976). Apparent duration of long meaningful events and meaningless intervals. *Memory & Cognition*, *4*, 215—220.

*Kowal, K. H. (1981). Growth of apparent duration: Effect of melodic and non-melodic tonal variation. *Perceptual and Motor Skills*, *52*, 803—817.

*Lass, N. J., & Conn, W. H. (1974). Time perception: A comparative study of time estimations in empty passive, speech-filled passive, and speech-filled active conditions. *Journal of Auditory Research*, *14*, 117—120.

Lejeune, H. (1998). Switching or gating? The attentional challenge in cognitive models of psychological time. *Behavioural Processes*, *44*, 127—145.

*Loehlin, J. C. (1959). The influence of different activities on the apparent length of time. *Psychological Monographs: General and Applied*, *73*, 1—21 Whole No. 474.

*Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory & Cognition*, *22*, 673—686.

*Marmaras, N., Vassilakis, P., & Dounias, G. (1995). Factors affecting accuracy of producing time intervals. *Perceptual and Motor Skills*, *80*, 1043—1056.

*Marshall, M. J., & Wilsoncroft, W. E. (1989). Time perception and the Stroop task. *Perceptual and Motor Skills*, *69*, 1159—1162.

*Martin, G. A., Shumate, M., & Frauenfelder, K. (1981). Experience of duration as a function of number of responses, task difficulty, and sex. *Perceptual and Motor Skills*, *53*, 139—145.

*Meyer, J., Shinar, D., Bitan, Y., & Leiser, D. (1996). Duration estimates and users' preferences in human–computer interaction. *Ergonomics*, *39*, 46—60.

*Miller, G. W., Hicks, R. E., & Willette, M. (1978). Effects of concurrent verbal rehearsal and temporal set upon judgments of temporal duration. *Acta Psychologica*, *42*, 173—179.

*Münzel, K., Gendner, G., Steinberg, R., & Raith, L. (1988). Time estimation of depressive patients: The influence of interval content. *European Archives of Psychiatry and Neurological Sciences*, *237*, 171—178.

*Neath, I., & Fortin, C. (2005). Is the interference between memory processing and timing specific to the use of verbal material? *Memory*, *13*, 395—402.

*Pedri, S., & Hesketh, B. (1993). Time perception: Effects of task speed and delay. *Perceptual and Motor Skills*, *76*, 599—608.

*Sawyer, T. F. (1999). Allocation of attention and practice in the production of time intervals. *Perceptual and Motor Skills*, *89*, 1047—1051.

*Schiffman, H. R., & Bobko, D. J. (1977). The role of number and familiarity of stimuli in the perception of brief temporal intervals. *American Journal of Psychology*, *90*, 85—93.

*Tsao, Y. -C., Wittlieb, E., Miller, B., & Wang, T. -G. (1983). Time estimation of a secondary event. *Perceptual and Motor Skills*, *57*, 1051—1055.

*Underwood, G., & Swain, R. A. (1973). Selectivity of attention and the perception of duration. *Perception*, *2*, 101—105.

*Vitulli, W. F. (2003). A systematic replication of variations in verbal content and delay on time estimation and short-term memory. *Perceptual and Motor Skills*, *96*, 1215—1222.

*Vitulli, W. F., & Shepard, H. A. (1996). Time estimation: Effects of cognitive task, presentation rate, and delay. *Perceptual and Motor Skills*, *83*, 1387—1394.

*Wilsoncroft, W. E., Stone, J. D., & Bagrash, F. M. (1978). Temporal estimates as a function of difficulty of mental arithmetic. *Perceptual and Motor Skills*, *46*, 1311—1317.

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

*Zakay, D. (1993b). Time estimation methods—Do they influence prospective duration estimates? *Perception*, *22*, 91—101.

*Zakay, D. (1998). Attention allocation policy influences prospective timing. *Psychonomic Bulletin & Review*, *5*, 114—118.

Zakay, D. (2000). Gating or switching? Gating is a better model of prospective timing (a response to 'switching or gating?' by Lejeune). *Behavioural Processes*, *50*, 1—7.

*Zakay, D., Nitzan, D., & Glicksohn, J. (1983). The influence of task difficulty and external tempo on subjective time estimation. *Perception & Psychophysics*, *34*, 451—456.

*Zakay, D., & Shub, J. (1998). Concurrent duration production as a workload measure. *Ergonomics*, *41*, 1115—1128.