



Symbolism overshadows the effect of physical size in supra-second temporal illusions

Hakan Karşılar^{1,2} · Fuat Balcı^{2,3}

Published online: 4 June 2019
© The Psychonomic Society, Inc. 2019

Abstract

The perception of quantities has been suggested to rely on shared, magnitude-based representational systems that preserve metric properties. As such, different quantifiable dimensions that can characterize any given stimulus (e.g., size, speed, or numerosity) have been shown to modulate the perceived duration of these stimuli—a finding that has been attributed to cross-modal interaction among the quantity representations. However, these results are typically based on the isolated effects of a single stimulus dimension, leaving their potential combined effects uncharted. In the present study we aimed to investigate the joint effects of numerical magnitude and physical size on perceived time. In four complementary experiments, participants categorized six durations as “short” or “long,” which were presented through combinations of Hindu–Arabic numerals in three font sizes, as well as with simple shapes (rectangles) and unfamiliar symbols (Klingon letters), the sizes of which corresponded to the font sizes of the Hindu–Arabic numerals. Our results showed temporal underestimation for the smallest numeral in the set (3), with no effects of font size on perceived duration. The perceived durations were longest for the physically smallest geometric stimuli (i.e., a rectangle), and the font size of symbol-like stimuli (i.e., Klingon letters) was not found to have an effect on perceived time. Finally, presenting only one numeral (6) instead of the rectangle once again eliminated the relationship between physical size and perceived time, suggesting an overshadowing of physical-size-based influences on temporal choice behavior, presumably by perceived symbolism. Our results point at the complex nature of the interaction between different magnitude representations.

Keywords Numerals · Size · Magnitude · Temporal bisection · Time perception

Accurate representation of the magnitudes of space, time, and numerosity is critical for organizing and integrating information (Allan, 1979; Buhusi & Meck, 2005; Meck, Penney, & Pouthas, 2008), which is fundamental for many forms of adaptive behavior, including foraging, hunting, and mating (see Gallistel, 1990; Gibbon & Church, 1990). Despite their seemingly unrelated physical counterparts, the representations of different quantities, such as spatial distances, time, and numerosity, have been argued to be functionally intertwined and to be underlain by overlapping neural systems (Buetti & Walsh, 2009; Conson, Cinque, Barbarulo, & Trojano, 2008;

Hubbard, Piazza, Pinel, & Dehaene, 2005; Pinel, Piazza, Le Bihan, & Dehaene, 2004; Walsh, 2003). This assertion of a common neurocognitive metric for different dimensions suggests that quantity judgments can interact with each other among disparate domains. In support of this prediction, larger, brighter, louder stimuli, or dynamic stimuli with faster motion, or stimuli at higher frequencies are perceived as lasting longer than stimuli with opposing features, suggesting functional overlap, particularly between the perception of physical and temporal magnitudes (e.g., Allan, 1984; Beckmann & Young, 2009; Brigner, 1986; Brown, 1995; Kaneko & Murakami, 2009; Karşılar, Kısa, & Balcı, 2018; Matthews, Stewart, & Wearden, 2011; Mo & Michalski, 1972). Moreover, a study by Oliveri et al. (2008) showed that these effects can generalize to symbols with well-learned semantic references, based on the finding that the corresponding magnitudes of Hindu–Arabic numerals can also affect temporal estimations (see also Dormal, Seron, & Pesenti, 2006; Oliveri, Koch, & Caltagirone, 2009; Xuan, Zhang, He, & Chen, 2007).

Specifically, Oliveri et al. (2008) found that numbers of higher value (e.g., 9) were perceived as lasting longer than

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13414-019-01748-x>) contains supplementary material, which is available to authorized users.

✉ Fuat Balcı
fbalcı@ku.edu.tr

¹ Department of Psychology, Ozyegin University, Istanbul, Turkey

² Department of Psychology, Koç University, Istanbul, Turkey

³ Koç University Center for Translational Medicine, Istanbul, Turkey

lower numbers (e.g., 5). The finding that participants are more accurate at classifying intervals represented by numerals as either “short” or “long” if the values of the numerals are congruent with the classified duration (e.g., a low number presented for a short duration; Xuan, Chen, He, & Zhang, 2009) supports an overlap between temporal and numerical processing. Similarly, a functional overlap between physical size and number processing has been proposed, in which participants can make faster decisions regarding the magnitude of a given numeral when the symbolic representation of a number matches its physical size (a small 3 or a large 7; Henik & Tzelgov, 1982; Kallai & Tzelgov, 2012). In turn, the processing of temporal magnitudes has been found to interfere with that of numerosity estimates (e.g., Brown, 1997). Finally, categorizations based on numerosity (i.e., “less/more”) were found to spontaneously transfer to durations, where the nature of this transfer function (at least in humans) has been argued to be based on the proportion of the compared magnitudes (Balci & Gallistel, 2006) instead of the raw magnitude representations, as had been previously thought (Meck & Church, 1983; Meck, Church, & Gibbon, 1985).

Although such results linking the representations of numbers, space, and time are abundant in the literature, no study so far has attempted to investigate the joint effect that the numerical and spatial dimensions can exert on subjective time. In this study, we attempted to elucidate this possible combined effect by utilizing various symbols of different font sizes with and without semantic content references, as well as simple geometrical shapes at corresponding sizes that are isolated from any possible attribution of symbolic meaning.

Although converging evidence has demonstrated cross-modal interaction and the transfer of information between different quantitative dimensions (see Alards-Tomalín, Leboe-McGowan, Shaw, & Leboe-McGowan, 2014), information-processing models aimed at explaining these interactions fall short of formally accounting for their representational overlaps, at both the neural and cognitive levels (but see Meck & Church, 1983; Meck et al., 1985). The most prominent of such models of temporal processing employs a pacemaker–accumulator theoretical approach, with three stages: (1) a pacemaker–accumulator component, which generates and temporarily stores durations in the form of pulses; (2) a memory component, in which the total number of pulses from the previous component is stored permanently; and (3) a decision/comparison stage, in which a random sample from the memory is compared to the value currently stored in the accumulator (e.g., “shorter/longer response”; Gibbon, Church, & Meck, 1984; see Allman, Teki, Griffiths, & Meck, 2014, for a review).

Although the basic schematics of this theoretical approach have been applied to other domains, as well (e.g., nonverbal counting; see Meck & Church, 1983), the pacemaker–accumulator family of models has garnered the largest amount of interest with regard to interval timing (see Grondin, 2010,

for a review). An internal-clock model (Treisman, 1963) utilizing such a multistaged information-processing delineation provides the analytical flexibility and tractability to account for distortions of time perception due to stimulus features (Coull, Vidal, Nazarian, & Macar, 2004; Penney, 2003; Wearden, 1999; Zakay & Block, 1995, 1997). This property allows for hypotheses to be postulated—for instance, in reference to physiological arousal (pacemaker), attention (switch as a temporal gating mechanism), information maintenance efficacy (accumulator), and/or decision biases (comparator)—while successfully accounting for well-established psychophysical properties (Balci & Simen, 2016; Simen, Rivest, Ludvig, Balci, & Killeen, 2013; Wearden & Lejeune, 2008).

Yet the theoretical underpinnings of such centralized internal-clock models generally do not readily explain how or why the magnitude information gathered/accumulated in one domain (e.g., numerosity or size) might influence magnitude judgments regarding another domain (e.g., duration). As such, existing pacemaker–accumulator models do not provide the necessary mechanistic ground for formulating clear directional hypotheses regarding the effect of other magnitudes on perceived duration. Nonetheless, these models provide the best parameter estimates for comparison among experimental conditions in which various audiovisual properties of stimuli are modulated—either individually or in unison—and the resultant warping of perceived time is measured. As such, they were broadly utilized in the present study in order to provide the generic theoretical grounds upon which the pattern of results could be discussed.

Furthermore, due to the metric properties of space, time, and numerosity (e.g., Montemayor & Balci, 2007), some researchers have attempted to reconcile these entities’ demonstrable interaction by adhering to the notion of a “common magnitude representation metric” (“a theory of magnitude”: *ATOM*; Walsh, 2003), allowing for the translation of magnitude codes among various quantitative dimensions. According to this approach, upon translation into a common magnitude code, any perceived magnitude becomes an approximate representation—that is, a unitless quantity (see also Balci & Gallistel, 2006)—and therefore becomes amenable to cross-modal comparison and transfer. Such an analog system utilizing a common neural metric also allows for cross-modal interaction, due to the uncertainty inherent in the neuronal information processing (Gallistel & Gelman, 2000). Arguments for such a common magnitude system have also received considerable support from neuroimaging research, which has consistently pointed to the parietal cortex—specifically, the intraparietal sulcus (Dormal & Pesenti, 2012)—as the neuro-anatomical locus for the processing of nonsymbolic as well as symbolic magnitudes (Bueti & Walsh, 2009; Hayashi et al., 2013). A natural extension of this proposed system includes subserving the perception of time, evidently in unison with other brain structures, such as the basal ganglia, frontal cortex,

and cerebellum (Apaydin et al., 2018; Fontes et al., 2016; Merchant, Harrington, & Meck, 2013). Although the assumptions of this common-magnitude-metric framework better accommodate many forms of cross-modal interactions reported in the literature, the framework does not make explicit predictions as to how any representation of a magnitude (e.g., a numeral) would interact in distorting subjective time. This is a shortcoming that the common-metric models (e.g., the ATOM model) share with the internal-clock models (e.g., STT), as we pointed out above. As such, it currently stands as a “stick-man” theoretical framework for fostering discussion on the subject of cross-modal interaction, rather than a tool through which directional hypotheses can be formulated and tested.

Although magnitude-based variables have been shown to share a directional relationship with perceived duration (larger size → longer duration; larger number → longer duration; Ono & Kawahara, 2007; Rammsayer & Verner, 2016; Xuan et al., 2007), no study so far has attempted to observe the combined effect of varying levels of both size and numerical value on perception of relatively long intervals. With the present study, we aimed to fill this empirical gap in the literature. Given the current literature, we expected there to be a congruity between the effects of font size and numerical magnitude, such that a 9 presented in a larger font would be perceived of as lasting longer than a 9 in a smaller font, which would, in turn, be perceived as lasting longer than 3 in the small font. Overall, our results in four complementary experiments were partially contradictory to this original hypothesis, showing that the mere implication of symbolism embedded within the timed stimulus can overshadow the effect of its physical magnitude (i.e., stimulus size; Exp. 1) on perceived duration. This investigation was achieved by utilizing simple shapes (e.g., a rectangle; Exp. 2) or meaningless symbols (i.e., Klingon letters; Exp. 3), along with Hindu–Arabic numerals (Exps. 1 and 4), as the timed stimuli in a commonly used time-based categorization task, namely the temporal bisection task (Church & Deluty, 1977). Interestingly, removing the inherent symbolism from the timed stimulus restored the effect of physical size on the timed intervals (i.e., smaller size led to overestimation), albeit in the opposite direction from that in the previous findings gathered with shorter time intervals. These results are discussed by adhering to prominent information-theoretic models of time perception.

Theoretical approach to temporal bisection

Although a number of modeling approaches have been proposed to account for temporal bisection data (primarily choice proportions), we will focus on one class of these, which has been proposed within the scalar expectancy/timing theory (SET) framework, in order to better demonstrate the

information-processing components of temporal bisection (e.g., Wearden & Ferrara, 1995). According to this specific approach, pulses generated by a Poisson process pacemaker are integrated within an “accumulator” during presentation of the test stimulus. The resultant collection of these pulses constitutes the real-time working memory representation of the elapsed duration. This representation, in turn, can be continuously compared to a random sample from the reference memory, containing a noisy representation of the arithmetic mean of the short and long reference durations. If the ratio of the difference between the total elapsed time (t) and the memory sample (M) to the total elapsed time [i.e., $(t - M)/t$] is below a given threshold (i.e., $-\beta$), the participant gives a short response; when it is above a given threshold (i.e., β), the participant gives a long response; and when it is in between these two thresholds, the participant gives a random categorization response (Wearden & Ferrara, 1995). Within this framework, if the speed of the pacemaker increases for any reason (e.g., due to a higher-than-baseline magnitude of the timing stimulus), the accumulated pulses would cross these thresholds earlier during the trial, resulting in a leftward shift in the resultant psychometric function, whereas a reduced speed of the pacemaker would result in a rightward shift in the psychometric function. Likewise, if the long-term memory of the target duration is disrupted (encoding a lower value than the actual working memory representation)—for instance, due to biases imposed during memory consolidation—the psychometric function would once again be expected to shift on the abscissa (e.g., in the form of a leftward shift in the case of lower-than-actual long-term memory representation of the duration; Levy, Namboodiri, & Hussain Shuler, 2015). Finally, with regard to the ease with which the durations are categorized, a shorter absolute distance between $-\beta$ and β (corresponding to a lower threshold when the integrated temporal evidence is $|t - M|/t$, thereby reducing the size of the guessing zone) and a lower memory variability would be expected to result in a steeper psychometric function, representing a smaller just-noticeable difference between the two reference durations. Likewise, a greater β distance and greater variability would result in a shallower function and a greater just-noticeable difference.

Experiment 1

Method

Participants

Thirty-four participants from Koç University (21 female, 13 male; $M_{\text{age}} = 20.1$ years, $\text{Range}_{\text{age}} = 18\text{--}29$) took part in the experiment. The sample size was based on the typical sample sizes used in similar studies (e.g., Rammsayer & Verner, 2016;

Xuan et al., 2007). Ten of the participants received one course credit, and 24 of the participants received 12 Turkish liras (approximately US \$4) for their participation. All participants had normal or corrected-to-normal vision. The study was approved by the Institutional Review Panel for Human Subjects of Koç University. All participants provided written consent for their participation.

Stimuli and apparatus

The stimuli consisted of three Hindu–Arabic numerals—3, 6, and 9—in different font sizes (small, medium, and large). The actual small, medium, and large font sizes presented on the test screen were approximately 2, 4, and 6 cm in height (69, 139, and 203 pixels: 1.91, 3.82, and 5.72 deg of visual angle, respectively) and 1, 2, and 3 cm wide (50, 99, and 144 pixels, respectively). All stimuli and instructions were presented on a 21-in. screen (60-Hz refresh rate) in a dimly lit room on an Apple iMac G4 computer. The stimuli were generated in Matlab using the PsychToolbox extension (Brainard, 1997; Pelli, 1997). The numbers 3, 6, and 9 in varying sizes were presented centrally in white on a gray background, with no text outline or borders (see Fig. 1 for a graphic depiction). Participants sat at a distance of approximately 60 cm from the screen and provided their responses using a mechanical, wired keyboard (Zalman ZM-K500).

Temporal bisection task

Training Each experimental session started with a training block, in which participants were first presented with the two *reference* durations (short and long: 1 and 3.5 s, respectively), represented by the presentation duration of circular black–white–gray mottled textures with a diameter of approximately 8 cm. The two reference durations were then presented five times each, in random order, and participants categorized the durations of the two stimuli as “short” or “long,” by

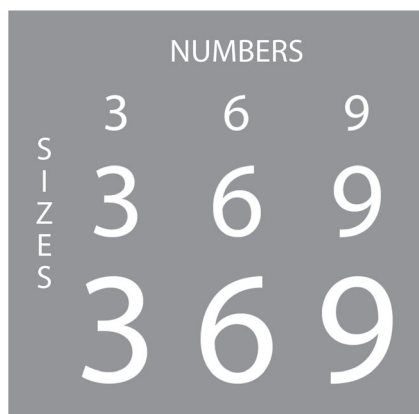


Fig. 1 Depiction of the relative sizes (small, medium, and large) of all numerals (3, 6, and 9) used in Experiment 1. Each of these numerals and size pairs served as timing stimulus signaling the duration to be judged

using one of two buttons on the keyboard (“F” or “K”). The key mappings were counterbalanced across participants. Participants were instructed not to count or to use any chronometric heuristics or methods throughout the experimental session (such as by keeping a rhythm). The training instructions explicitly stated that only one of the two previously experienced durations, and no intermittent ones, were being presented at this stage of the experiment. A 1-s feedback (“Correct”/“Incorrect”) was given visually upon response in the training trials. An incorrect response was followed by a correction trial with identical parameters. When participants had achieved a total of ten correct responses, the training block was terminated.

Test For the remainder of the session, participants’ task was to categorize (as “short” or “long”) six probe durations (1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 s), represented by the presentation duration of one of the three numerals (3, 6, or 9) in one of three font sizes (small, medium, or large; Fig. 1). The key mapping for response collection was identical to that used in the training session. All possible combinations of the two variables (3×3) randomly appeared ten times for all probe durations, leading to a total of 540 trials per participant. Each stimulus was presented after a participant pressed the “space” button on the keyboard, followed by a fixed stimulus-to-response-prompt interval of 0.5 s, after which participants were allowed to respond. No feedback was given to participants as to the accuracy of their choices. Participants took part in a single 50- to 60-min session and were instructed to take a break any time they felt tired or failed to maintain their attention on the task. A compulsory break of 30 s was given automatically every 5 min throughout the session.

Data analysis

The data were initially analyzed with frequentist factorial repeated measures analyses of variance (ANOVAs) and their corresponding post-hoc tests. In addition to these conventional analyses, we also report the results of their Bayesian counterparts (i.e., Bayesian ANOVA and Bayesian paired *t* tests), for which we report inverse Bayes factors (BF_{10} : i.e., the strength of evidence that the data provide for the alternative as compared to the null hypothesis; see, e.g., Wagenmakers, Marsman, et al., 2018), allowing for easy comparison of parameters (see Held & Ott, 2018). Values of BF_{10} in the intervals 1–10, 10–100, and 100–300 are interpreted as providing weak to moderate, moderate to strong, and strong to decisive evidence for the alternative hypothesis, respectively, whereas the inverses of these values ($1/BF_{10}$) provide evidence for the null hypothesis in identical descriptive intervals (BF_{01} ; Goodman, 1999; Jeffreys, 1961). Hence, values of BF_{10} less than 1 provide support for the null hypothesis. We used the

JASP 9.0.2 open-source software with default priors for all the Bayesian tests (see JASP Team, 2018; Wagenmakers, Love, et al., 2018). In all of the Results sections below, we indicate the model that has the highest Bayes factor with respect to the null model. When testing the interaction between variables, the likelihood of the model that contained the main effects was compared to that of the model containing the main effects as well as the interaction.

Results

For each combination of the three levels of the independent variables (numeral and font size), a total of nine sigmoidal psychometric functions were formed for each participant by plotting the mean percentage of “long” responses as a function of the six probe durations (see Figs. 2A and B for psychometric functions fit to data averaged across participants). A cumulative Weibull distribution function was fit to each of these plotted data, and the point of subjective equality (PSE) was calculated as the median of the best-fit Weibull function ($R^2_{\text{Mean}} = .946$, $R^2_{SD} = .045$), using custom scripts written in Matlab (R2018B, The MathWorks, Inc.). The PSE is known as the duration at which “short” and “long” responses are equiprobable. In other

words, PSE is the duration at which the participant would theoretically find it most difficult to distinguish between the “short” and “long” options, and is therefore equally likely to pick either one. This measure can be used to detect any changes in the speed of an internal clock (i.e., increased speed leading to leftward shifts in the PSE). We were primarily interested in the modulation of the PSE values as a function of numerical value, font size, or a combination of the two. Additionally, as a measure of the *steepness* of the Weibull function indicating the discriminability of the durations, Weber ratios (WRs) were calculated by dividing the difference limen [DL: $(p(\text{long}) = .75 - p(\text{long}) = .25)/2$] by the PSE. Along with PSE values, the WR values were used as dependent measures in the analyses detailed below. The parameters of Weibull fits with adjusted- R -squared values less than .7 were replaced by a random value drawn from the sample distribution so that they did not alter the mean or the standard deviation of the sample (1.4% of the total cases). This replacement procedure was carried out so that the entire dataset of a given participant with a bad Weibull fit for a single experimental condition was not excluded from the overall repeated measures ANOVA procedure. Two participants with more than two such “bad” Weibull fits across all nine conditions were excluded from further analyses in

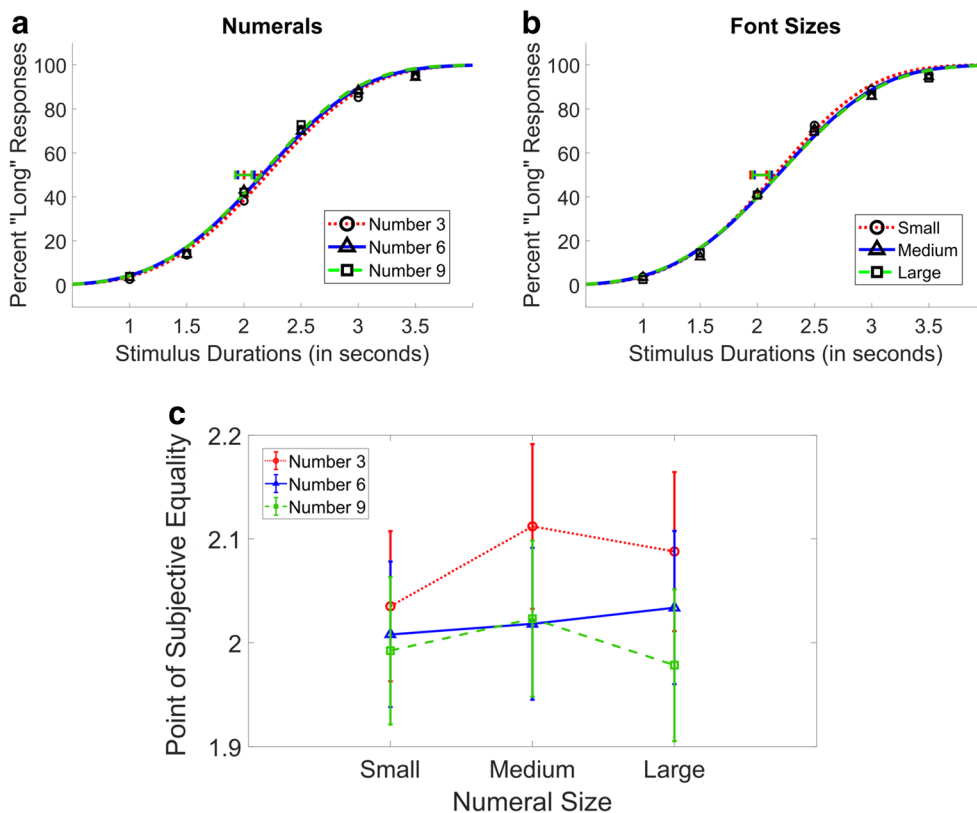


Fig. 2 Weibull functions fit to the mean percentages of “long” responses as a function of probe durations in Experiment 1 for the three levels of the (A) numerical value and (B) font size variables. Standard errors of the means of individually calculated points of subjective equality (PSEs) are marked with horizontal lines of colors identical to those in the different

conditions. (C) PSE values (.5 likelihood of reporting a duration as “long,” in seconds) calculated from the Weibull fits for combinations of the font size and numeral variables. Error bars denote within-subjects errors (95% confidence intervals [CIs]; Cousineau, 2005) based on the number of participants (Masson & Loftus, 2003)

Experiment 1. This a priori exclusion criterion was also employed in all further experiments (see below).

A two-way repeated measures ANOVA was conducted with PSE as the dependent variable (higher PSEs representing lower temporal integration/shorter perceived durations) and numerical value (three levels) and font size (three levels) as within-subjects factors (see Fig. 2C). The results showed a main effect of numerical magnitude [$F(2, 62) = 6.41, p = .005, \eta_p^2 = .167$], and no main effect of font size [$F(2, 62) = 0.42, p = .4, \eta_p^2 = .043$] or interaction between the two variables [$F(2.98, 92.48) = 1.11, p = .4, \eta_p^2 = .022$, Greenhouse–Geisser-corrected]. Post-hoc analyses showed that presentation of the number 3 led to the highest PSE ($M = 2.08$), as compared to the PSEs for the numbers 6 and 9 ($M = 2.01, p = .041, d = 0.38$, and $M = 2.01, p < .001, d = 0.72$, respectively, Holm–Bonferroni-corrected), whereas the difference between the latter two numerals was not significant ($p = .4, d = 0.15$). The Bayesian two-way ANOVA with the same variables revealed an identical pattern of results, in which the data provided strong evidence for the model that included the effect of numerical magnitude over the null ($BF_{10} = 10.81$). A two-way repeated measures ANOVA with identical within-subjects factors and WR as the dependent variable revealed no main or interaction effects ($p_{\text{number}} = .08, p_{\text{size}} = .65, p_{\text{interaction}} = .69$), and the best Bayesian model with the numerical magnitude variable failed to provide evidence for an effect on WR values ($BF_{10} = 0.61$).

Interim discussion

The results from Experiment 1 went against the initial hypothesis that numerical and physical magnitude would exert a combined effect (linearly combined or otherwise) on perceived duration. Although we did observe an effect of symbolic magnitude, in which the numeral 3 was perceived to last for a shorter duration than the numerals 6 and 9, the expected effect of font size (i.e., physical magnitude) was not observed. As such, in order to isolate the symbolic component of the observed effect from the effect of physical magnitude, further experiments commenced. Isolating and thereby entirely removing the symbolic representation of Hindu–Arabic numerals necessitated the use of nonsymbolic stimuli that would be as similar as possible to those that had been timed in Experiment 1. To test the effect of physical size itself on perceived time, an additional experiment was devised in which any symbolic *implication*, familiar or not, was completely removed from the stimuli, whereas the size manipulation was preserved. As such, varying sizes of simple rectangle shapes were

introduced into the temporal bisection procedure in Experiment 2.

Experiment 2

Method

Participants

Thirty-four participants (22 female, 12 male; $M_{\text{age}} = 20$ years, $\text{Range}_{\text{age}} = 18\text{--}24$) took part in the experiment for course credit. The study was approved by the Institutional Review Panel for Human Subjects of Koç University. All participants provided written consent for their participation.

Stimuli, apparatus, and procedure

The apparatus and training/test procedures employed in Experiment 2 were identical to those in Experiment 1. The stimuli used in Experiment 1 (the numerals 3, 6, and 9) were replaced by white-filled rectangles (thereby reducing the number of factors from two to one). Three sizes of rectangles were used in the experiment, in which the width and the height of the filled area corresponded to the limits of boxes encompassing the three font sizes used in Experiment 1 (i.e., each rectangle was treated as if it were also a letter with three font sizes). The length of the session and the number of trials were as in Experiment 1; thus, three times more data per condition were collected in Experiment 2.

Results

PSE values were calculated for each size of the rectangle stimulus by fitting cumulative Weibull functions to participants' data ($R^2_{\text{Mean}} = .974, R^2_{SD} = .02$; Fig. 3A). The parameters of Weibull fits with adjusted-*R*-squared values less than .7 were replaced by random values drawn from the sample distribution (1% of total cases). Two participants with more than two bad Weibull fits were excluded from further analyses.

A one-way repeated measures ANOVA with PSE as the dependent variable and rectangle size (three levels) as the within-subjects variable revealed a significant effect of rectangle size on perceived duration [$F(2, 62) = 4.53, p = .015, \eta_p^2 = .13$]. Post-hoc simple-effect analyses showed that the PSE in the small rectangle-size condition ($M = 1.99$) was significantly lower than those in the medium ($M = 2.05, p = .02, d = 0.5$, Holm–Bonferroni-corrected) and large ($M = 2.04, p = .045, d = 0.41$, Holm–Bonferroni-corrected) rectangle-size conditions, whereas the difference between the latter two did not reach significance ($p = .65, d = 0.08$, Holm–Bonferroni-corrected; see Fig. 3B). This result was supported by the Bayesian analysis, which provided moderate support for the

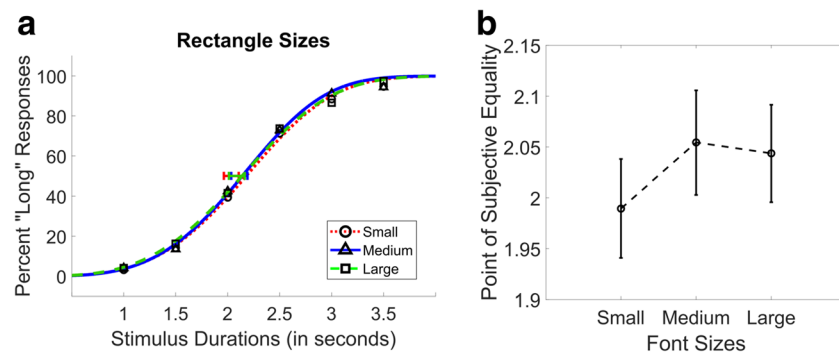


Fig. 3 (A) Weibull functions fit to the mean percentages of “long” responses as a function of probe duration in Experiment 2, for the three stimulus (rectangle) sizes. (B) PSE values (in seconds) calculated for the

three rectangle sizes. Error bars denote within-subjects errors (95% CIs). All other aspects of the figure are as in Fig. 2

alternative over the null hypothesis ($BF_{10} = 3.04$). Finally, a one-way repeated measures ANOVA with the same independent variable and WR as the dependent variable revealed no difference between the three conditions [$F(2, 62) = 0.99, p = .38, \eta_p^2 = .031$], and an identical Bayesian analysis provided no evidence for the effect of size over the null model with regard to WR ($BF_{10} = 0.2$).

To confirm that the results above were not simply due to three times the data being collected in Experiment 2 as in Experiment 1, identical analyses were conducted on batches of data that were randomly selected for each participant such that the number of trials per condition decreased by a factor of three. These results showed a virtually identical pattern of frequentist significance of the model parameters. The patterns of Bayesian support for the null or the alternative hypotheses were also identical.

Interim discussion

The results from Experiment 2 showed that, when all forms of symbolic meaning (implied or known) are experimentally removed, the physical size of stimuli can indeed modulate the perceived duration of the stimulus, so that the smallest stimulus leads to the longest perceived duration. Interestingly, the direction of our results opposes those from the relevant literature, in which smaller objects are expected to slow down temporal integration, thereby constricting perceived intervals. Since we observed no effect of font size in Experiment 1, the results of Experiment 2 suggest the possibility that symbolic meaning overshadows the effect of size on perceived time. Testing this hypothesis necessitated the isolation and removal of symbolic meaning (i.e., 9 implying a numerosity of nine objects, which is semantically larger than 3) from the timed stimuli, while preserving their symbol-like features. Therefore, in a follow-up experiment, we devised a test in which the isolated effect of symbolism without numerical value was tested in identical conditions with stimuli that appeared to carry alphabetical/numerical information, yet that had no

specific meaning to the observer. For this purpose, we utilized Klingon letters as the to-be-timed stimuli in Experiment 3, since they have the appearance of carrying symbolic meaning similar to that of the Hindu–Arabic numerals, yet that meaning would be alien to any participant who had no previous association with the Star Trek or familiarity with the said fictional alphabet (see Fig. 4).

Experiment 3

Method

Participants

Thirty-six participants (20 female, 16 male; $M_{\text{age}} = 20.2$ years, $\text{Range}_{\text{age}} = 18\text{--}28$) took part in the experiment for course credit. One male participant was excluded from the study, due to high familiarity with the Star Trek franchise and the associated Klingon alphabet (Okrand, 1992; see the Stimuli, Apparatus, and Procedure section below). All remaining participants had normal or corrected-to-normal vision and reported little to no knowledge or familiarity with the Star Trek franchise or the Klingon alphabet. The study was approved by the Institutional Review Panel for Human Subjects of Koc



Fig. 4 Depiction of the three Klingon letters used in Experiment 3. These letters correspond to lowercase letters “l,” “m,” and “p,” from left to right (Okrand, 1992). The typeface was provided by and printed with permission from the Klingon Language Institute

University. All participants provided written consent for their participation in the experiment.

Stimuli, apparatus, and procedure

The apparatus and training/test procedures employed in Experiment 3 were identical to those in Experiment 1. The stimuli used in Experiment 1 (the numerals 3, 6, and 9) were replaced by three Klingon letters quasi-randomly selected by the authors with subjective normative criteria, originally corresponding to lowercase letters “l,” “m,” and “p” of the KLI pIqaD script (see Fig. 4). Although some participants reported that the first letter resembled the capital “L” upon debriefing (see Fig. 4, leftmost symbol), none of the participants reported any perceived resemblance of the timed stimuli to a known numerical system, whether Hindu–Arabic or otherwise. The font sizes of these letters were also identical to those used in Experiment 1 (see Fig. 1 for reference).

Results

As in the preanalysis procedure used in Experiment 1 above, PSE values were calculated for each possible

combination of letter and font size by fitting cumulative Weibull distribution functions to the “long” choice proportions, plotted against the six probe durations ($R^2_{\text{Mean}} = .937$, $R^2_{SD} = .047$; Presented as percentages in Figs. 5A and B). The parameters of Weibull fits with adjusted- R -squared values less than .7 were replaced by random values drawn from the sample distribution (2.4% of total cases). Three participants with more than two bad Weibull fits were excluded from further analyses.

A two-way repeated measures ANOVA was conducted, with PSE as the dependent variable and symbol (three levels) and font size (three levels) as within-subjects factors (Fig. 5C). The results showed no main effect of symbol type [$F(2, 64) = 0.94$, $p = .4$, $\eta_p^2 = .028$] or main effect of font size [$F(2, 64) = 0.66$, $p = .52$, $\eta_p^2 = .02$], and there was no interaction between the two factors, either [$F(4, 128) = 1.06$, $p = .38$, $\eta_p^2 = .032$]. This finding was confirmed by the Bayesian analysis with identical variables, in which the null model was preferred over the best model explaining the data with the Klingon symbol variable ($BF_{10} = 0.09$). A two-way repeated measures ANOVA with identical within-subjects factors and WR as the dependent variable also revealed no main or interaction effects ($p_{\text{number}} = .51$,

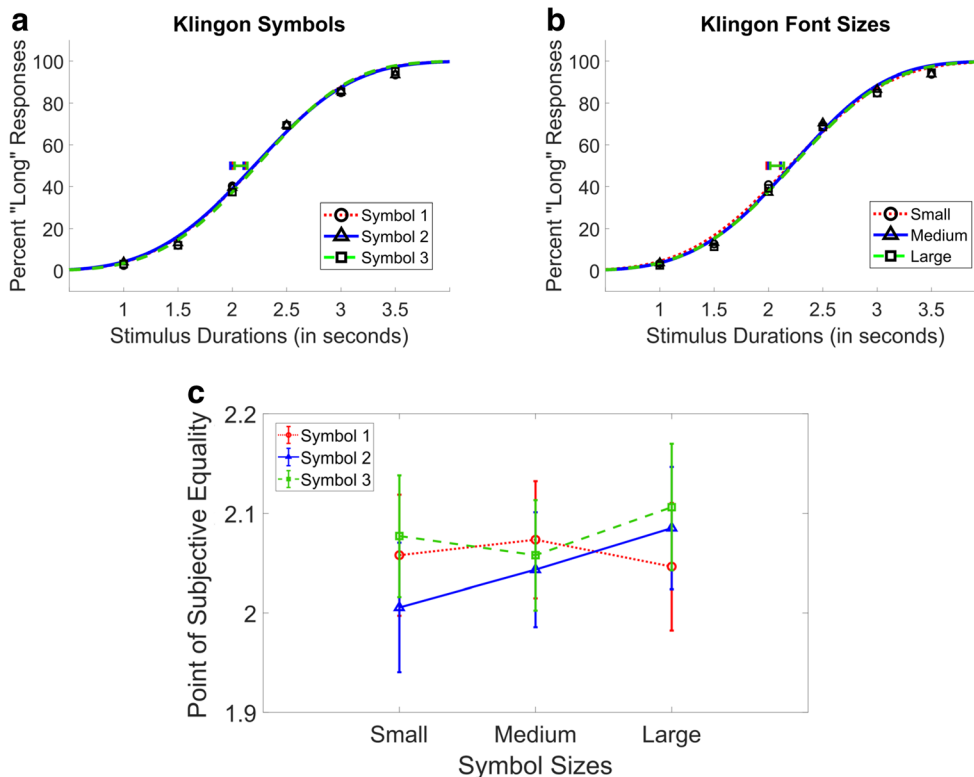


Fig. 5 Weibull functions fit to the mean percentages of “long” responses as a function of probe duration in Experiment 3, for the three levels of the (A) Klingon symbol and (B) font size variables. (C) PSE values (.5

likelihood of reporting a duration as “long,” in seconds) calculated from the Weibull fits for combinations of the Klingon symbol and font size variables. All other aspects of the figure are as in Fig. 2

$p_{\text{size}} = .58$, $p_{\text{interaction}} = .38$), and the best Bayesian model with the Klingon symbol variable failed to provide evidence for an effect on WR value ($BF_{10} = 0.08$).

Interim discussion

The results from Experiment 3 suggested that, as expected, when the symbolic meaning was removed, the symbols did not exert any effect on perceived durations, as opposed to, say, the number 3 being timed shorter than numbers 6 and 9 in Experiment 1. Interestingly, even though any symbolic meaning was removed from the stimuli, Experiment 3 also failed to reveal an effect of font size on perceived duration, as had been observed in Experiment 2. Taken together with the results of Experiment 2, in which an effect of physical size was demonstrated when the stimuli had no implied symbolism, these results suggest that physical size indeed cannot exert an effect on perceived durations when there is merely the implication of symbolism to the timing stimulus. At this point with regard to the working hypothesis of this study, the procedural differences between Experiments 1 and 3 and Experiment 2 remained to be addressed. Namely, while the design of Experiment 2 had eliminated all symbolically exerted effects on perceived time, it simultaneously introduced a procedural nonparallelism with Experiments 1 and 3, such that dropping one of the factors (symbolism) from the experiment by replacing the stimuli with a single geometrical shape meant that the participants were concerned with timing stimuli that varied on only one dimension (size) throughout the session—in other words, half the dimensions of the previous experiments. In turn, it is possible that the salience of the dimension on which the participants observed any systematic change (physical size) might have acted as a confound in our experimental design. As such, in order to conclusively suggest that our results were directly related to the effect of change in physical size and were, in fact, independent of the effect of observing changes on only one dimension (as opposed to two, in Exps. 1 and 3), an additional experiment was devised in which the timed stimuli were different sizes of one, single Hindu–Arabic numeral.

Experiment 4

Participants

Thirty-six participants (21 female, 15 male; $M_{\text{age}} = 20.9$ years, $\text{Range}_{\text{age}} = 18\text{--}30$) took part in the experiment for course credit. The study was approved by the Institutional Review Panel for Human Subjects of Koç University. All participants provided written consent for their participation in the experiment.

Stimuli, apparatus, and procedure

The apparatus and training/test procedures employed in Experiment 4 were identical to those in Experiment 2, with the exception that the stimuli used were changed from a rectangle to the number 6 that had been used in Experiment 1 (see Fig. 1). The three font sizes used in Experiments 1, 2, and 3 were also used in this experiment. The length of the session and the total number of trials were identical to those aspects of Experiments 1, 2, and 3; thus, as in Experiment 2, three times more data per stimulus type were collected in Experiment 4 than in Experiments 1 and 3.

Results

PSE values were calculated for each font size of the number 6 stimulus by fitting cumulative Weibull functions to participants' data ($R^2_{\text{Mean}} = .963$, $R^2_{SD} = .024$; Fig. 6A). The parameters of Weibull fits with adjusted- R -squared values less than .7 were replaced by random values drawn from the sample distribution (1% of total cases). Three participants with more than two bad Weibull fits were excluded from further analyses.

A one-way repeated measures ANOVA with PSE as the dependent variable and font size (three levels) as the within-subjects variable failed to reveal a significant difference in PSEs among the three font sizes [$F(2, 64) = 2.63$, $p = .08$, $\eta_p^2 = .076$, Fig. 6B], a finding confirmed by the Bayesian analysis, which showed preference for the null over the alternative hypothesis ($BF_{10} = 0.73$). Similarly, an identical analysis with WR as the dependent variable revealed no difference between the three conditions [$F(2, 64) = 0.26$, $p = .77$, $\eta_p^2 = .008$], and an identical Bayesian analysis provided no evidence for an effect of size over the null model with regard to WR ($BF_{10} = 0.11$).

General discussion

Physical magnitudes in different domains tend to be correlated. For instance, a larger rock tends to generate more noise as it rolls down a cliff. Such relations might also be inherently present in the cognitive representation of spatio-temporal information, as a result of lifelong learning or due to the evolved architecture of the magnitude perception system. In light of convergent behavioral evidence demonstrating an interaction among the perceptions of different physical domains (see Eagleman, 2008, for a review), a theoretical framework for a translational system for analog representations has been postulated (Walsh, 2003). Within this framework, representations of space, time, and numerosity are neurally coded as unit-free

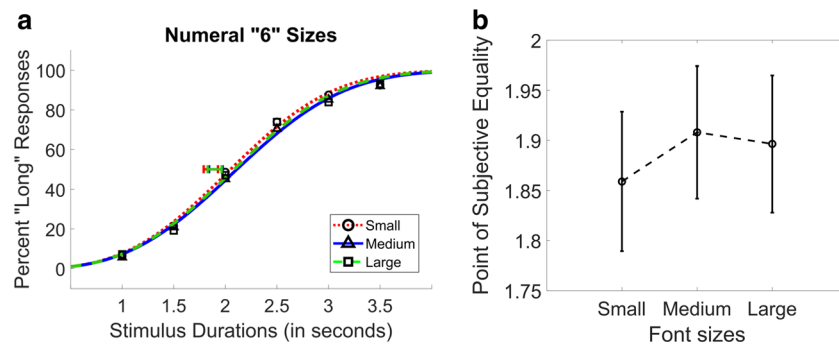


Fig. 6 (A) Weibull functions fit to the mean percentage of “long” responses as a function of probe duration in Experiment 4, for the three font (number 6) sizes. (B) PSE values (in seconds) calculated for the three

font sizes. Error bars denote within-subjects errors (95% CIs). All other aspects of the figure are as in Fig. 2

quantities (i.e., “a little” or “a lot”) instead of in terms of their absolute magnitudes (see also Balci & Gallistel, 2006, for the use of proportions as the metric), thereby allowing for interactions within this “noisy” (see Libertus, Feigenson, & Halberda, 2013; Meck & Church 1983), all-purpose magnitude transactional system (Buetti & Walsh, 2009). However, empirical studies on the interaction among magnitude representations have tended to select only two domains within which the effect is demonstrated of an experimental manipulation in one domain (e.g., presenting a larger disk or a larger number) on judgments in the other (e.g., dilation of subjective time; Rammsayer & Verner, 2016; Xuan et al., 2007).

Although these types of designs may be procedurally desirable, since they represent a straightforward and simplified approach to the question at hand, they nevertheless fail to capture the very process of integration of simultaneously extracted information from multiple domains and the interplay between them, a matter directly related to our everyday experience. Additionally, previous studies have generally opted to utilize subsecond test intervals (e.g., Ono & Kawahara, 2007; Thomas & Cantor, 1976; Xuan et al., 2007), possibly motivated by an attempt to elucidate the underlying dynamics of low-level processes directly related to our timing ability, such as motor preparation, speech production, and fast actions. Yet, similar to the problems inherent in using perceptually simplified stimuli that only vary on a single dimension, our everyday experience related to timing also takes place in the supra-second (i.e., “longer than 1 s”) durations, which necessitates exploring behavior within these ranges, as well. As such, given the theoretical assertions suggesting that numerals gain meaning by being mapped onto magnitude representations of numerosities (e.g., Gallistel & Gelman, 1992, 2000), we have extended and added to this questioning in the literature in order to encompass the symbolic representation of magnitude information and their physical features, by devising an experimental procedure in which the numerical magnitude and physical size of the timing stimuli (i.e., Hindu–Arabic numerals) could

be modulated simultaneously. Thus, this test is also important for testing the generalizability of the previously reported effects to symbolic numerals and to longer time intervals.

In line with the previous literature, our results suggest that the presentation of larger numerical values indeed lengthened perceived durations (Exp. 1). However, in contradiction with the literature, the presentation of physically smaller or larger numerals (in terms of font size) had no discernible effect on subjective time. Hence, this otherwise effective predictor of temporal modulation (e.g., Ono & Kawahara, 2007; Rammsayer & Verner, 2016; Xuan et al., 2007) failed to show an effect on perceived time when the timed stimuli were Hindu–Arabic numerals (i.e., bearing symbolic meaning). In Experiment 2, we replaced the symbol-like stimuli with one simple geometric shape (a rectangle), which led to the modulation of perceived time with stimulus size. However, interestingly and again contrary to the literature (Ono & Kawahara, 2007; Xuan et al., 2007), we found that a smaller stimulus size led to the lengthening of perceived time (a leftward shift in PSE). In Experiment 3, stimulus size once again failed to modulate perceived time, even when the timed stimuli were composed of letter-like symbols (i.e., Klingon letters). Finally, we reasoned that the differential effects of stimulus size on perceived time between Experiment 2 and Experiments 1 and 3 could have been due to the fact that the stimulus sets were composed of three different stimuli in Experiments 1 and 3 (three different numeral/letter/numeral-like symbols), but only a single stimulus (i.e., rectangle) in Experiment 2. In order to control for this potential confound, we conducted a fourth experiment in which a single Hindu–Arabic numeral was used to test the effect of stimulus size on perceived time. The results of Experiments 1 and 3 suggested that the implication of symbolism, whether meaningful or not, can overshadow the effect of stimulus size on perceived time. In light of the contrasting results gathered from Experiments 2 and 4, and in conjunction with the results from Experiments 1 and 3, we concluded that there was no effect of stimulus size when the timing stimulus was a numeral once again (despite the much larger number of trials per

stimulus in Exp. 4). Importantly, for those results with evidence of a significant effect, as it stands, the effect size of numerical magnitude in Experiment 1 is comparable to those from other studies (e.g., Rammsayer et al., 2016) but the size of the effect of physical size in Experiment 2 is lower than those generally found in the literature. This difference may have emerged for a number of reasons, including differences in stimulus properties and experimental methodology.

Under the umbrella of the term “size–number congruity effect,” it has been shown that participants can process numerical magnitude and numerical size simultaneously (e.g., Henik & Tzelgov, 1982). As we stated above, other studies have shown an interaction between the size of a stimulus and its perceived duration (Rammsayer & Verner, 2016), and completing the interactive triad of “time–number space,” participants’ judgments of durations have been shown to be affected by both numerosity (Dormal, Seron, & Pesenti, 2006; Xuan et al., 2007) and the symbolized value of the timed stimulus (Oliveri et al., 2008; Vicario, 2011). To the best of our knowledge, the present study is the first one to test all components of this triad of dimensions simultaneously, and the lack of replication of previous findings should provide us with valuable insight into the underlying dynamics of the interaction between space, time, and number (see Walsh, 2003). On that note, when taken as a whole, our results suggest that the magnitude information conveyed by their symbolic references is sufficient to induce clock-speed-like effects, whereas the physical size of these symbolic references does not lead to the rather well-established effects of the stimulus dimensions on time perception (also see Exp. 2). This finding is contrary to what could be hypothesized given the current literature, since it should be more likely for size to override numerical magnitude in a timing task such as the present one, given that interference from a continuous dimension such as size tends to be stronger than interference from discrete dimensions such as numerosity (Hurewitz, Gelman, & Schnitzer, 2006; but note that these researchers did not use numerals but sets of different sizes.) We find this an interesting and important empirical observation, because it adheres to the two postulates of symbolic representations: (a) Symbols are linked to semantic content and activate it (Gallistel & Gelman, 1992), and (b) symbols are arbitrary, with no inherent resemblance to their semantic content (e.g., 3 in the Hindu–Arabic numeral system and 11 in binary code refer to the same quantity). Consequently, one would not expect the physical size of a symbol to tap into a magnitude-representational interaction (Piazza, Pinel, Le Bihan, & Dehaene, 2007), because it is irrelevant to the semantic referencing by symbols. In other words, the arbitrariness of the symbol–semantics relation would be expected to overshadow the induction of a stimulus size–duration interaction in our specific experimental design. Interestingly, if such an “overshadowing effect” is

indeed the correct inference to be drawn from the results of our first experiment, our third experiment also adds to this situation the fact that the semantic component of the symbolic stimuli need not be actual, but in fact might only be *implied*, through symbols with no inherent connection to actual semantic content. Although we do not have an architectural account for the overshadowing effect of symbol-like characters, one of the possibilities is that symbol-like characters attract more attention to their form than to their physical properties, given the irrelevance of the latter for what and how the symbol-like character indexes information.

It is, however, also possible that—at least in special cases in which the informational value and the ease and automaticity with which magnitude information is derived from one domain (i.e., numerical value) is significantly higher than in another (i.e., size), as was the case in Experiment 1—the common magnitude representational system might actively dampen the information flow from the latter, more unreliable source of information. This is similar to linguistic category information becoming a better predictor of color judgments when more fine-grained perceptual information suffers from relatively higher uncertainty (Cibelli, Xu, Austerweil, Griffiths, & Regier, 2016). Thus, our empirical observations in Experiment 1 might not be peculiar to symbolic representations; well-learned and widely used symbols might instead simply meet the requirements of high stimulus saliency. Nonetheless, our findings in Experiment 3 cast some doubt on this claim, since those stimuli fundamentally lacked learning and familiarity. However, all things considered, it seems likely that the underlying hierarchy of the magnitude representation system (whether in an interactive manner or not) dictates the prioritization of one form of magnitude information over others—in this case, symbolic meaning over physical inference—pointing to a fundamental asymmetry of the time–number space interaction in the brain (see also Martin, Wiener, & van Wassenhove, 2017). Explicating at least one side of this asymmetry, theories of time perception have included among their ranks those that propose the perception of time to be a by-product of any and all other cognitive processes (Karmarkar & Buonomano, 2007; Muller & Nobre, 2014), putting the perception of time in a unique condition. Thus, it is possible that although our perception of time might interact with individual dimensions of magnitude perception, this interaction need not be one whose metrics and cognitive mechanisms are constant across all modalities, especially when they vary in terms of their salience and informational content for the specific task at hand.

On another note, our differential findings might have more to do with innate architectural rather than situational factors; the “cross-modal affinity” of magnitude-based representations might vary for different domains. For instance, as compared to spatial manipulations, it might by default be easier to induce effects of numerosity-related manipulations on

duration judgments, due to the inherent properties of specific magnitude-based representations and/or a higher degree of overlap between their corresponding neural substrates. This might result in the a priori prioritization of certain kinds of information over others, particularly if the channels of cross-modal interaction/integration are constrained by the limits of cognitive resources such as attention (e.g., Kahneman, 1973). These possibilities constitute a fertile theoretical ground for future studies on interval timing specifically, and on magnitude representations in general.

Surprisingly, the effect of stimulus size on perceived time in the present study was in the opposite direction from those previously reported in the literature. As was stated above, many studies have investigated the effect of nontemporal magnitudes on interval timing for relatively short, typically subsecond intervals. The mismatch between our findings based on supra-second intervals and earlier findings based on subsecond intervals might be informative regarding the nature of cross-modal interactions with time and the function of these interactions. For instance, the often reported cross-modal interaction in the shorter interval range (i.e., tens of milliseconds to a second) could serve the function of guiding fast motor plans of the organism to meet the quantitative demands of a given environment, whereas for long intervals, these magnitudes might interact through the differential allocation of attentional resources between different quantities, thereby affecting perceived duration differently than in the subsecond range. Although this supposition is rather speculative as of the writing of this report, less attentional resources would be expected to be devoted to smaller nontemporal magnitudes within this approach, leaving more attentional resources for the processing of time, and thus resulting in the dilation of perceived time (Penney, 2003; Zakay & Block, 1996, 1997). Another possible source of discrepancy between the results of the present study and previous work was the different tasks used. Previous tasks that had shown effects of stimulus size on time perception had used temporal reproduction (e.g., Rammsayer & Verner, 2016) and temporal discrimination (Xuan et al., 2007) procedures, whereas in the present study we used the temporal bisection task. Different tasks are known to lead to differential results in different contexts (e.g., Gil & Droit-Volet, 2011; Kroger-Costa, Machado, & Santos, 2013). Future studies will be needed to elucidate the potential reasons that underlie these discrepancies.

Finally, whereas the effect of numerals on perceived time adhered to those from the previous literature in Experiment 1 (Oliveri et al., 2008), it still remains to be explained why we found a significant difference in timing between the number 3 and the numbers 6 and 9, while no difference was found between the numbers 6 and 9 themselves. A potential explanation of our findings relates to the use frequency and the overall symbolic familiarity of different numerals in daily life. Specifically, the number 3 appears to carry with it a unique

affinity to human behavior, which is not shared by the numbers 6 and 9. There are two empirical bases for this speculation. First, it has been shown that the “preference” for the number 3 in daily life is significantly higher than other numerals, second only to the number 7 in subjective rankings (Bellos, 2015). Second, according to Newcomb–Benford’s law (Berger & Hill, 2011), the frequencies of the numbers 6 and 9 as leading numerals in naturally occurring statistical phenomena are exponentially lower than those of smaller numbers, including the number 3, which arguably is reflected in the differential usage and therefore of the symbolic representation of this specific number in daily settings. Overall, it can be argued that over the course of its presentation, the nontemporal properties of the numeral 3 may have grabbed more attention than the other numerals, leading to a significant underestimation of its presentation duration on the screen. This effect could have been coupled with the relative discriminability of the numerals used in this study. Namely, since we linearly scaled the numbers used in the experiment, the discriminability of 6 and 9 was lower than the discriminability of 3 and 6, and such differences are known to be manifested in the processing of numerals (Moyer & Landauer, 1967). These factors put together might have singled out the trials in which the numeral 3 was used in this study, and thus its differential processing during time perception, whereas the numbers 6 and 9 were treated almost equally. It should be noted, however, that the primary discussion of the results above (suggesting that symbolic implication could be sufficient to overshadow the effects of less readily processed magnitudes such as physical size on perceived time) would need to be restructured so as to confine this effect to include specific numbers with high cultural relevance/preference over others. A natural suggestion to test this explanation would be to utilize logarithmically scaled intervals in future studies (e.g., 1, 3, and 9), along with designs that distinguish between numbers with high versus low relevance to humans (e.g., 1, 2, and 6), while ensuring that single digits are used throughout the study, in order to avoid potential confounds introduced by multiple digit processing.

In conclusion, although our results are novel and highly suggestive of an asymmetry between the interaction of disparate magnitudes on perceived time, future studies will be needed in order to parse out the possible interaction between the size of the stimulus set and stimulus features in terms of their effects on interval timing. One drawback of our study is that numbers are automatically processed and are fairly absolute, but without any contextual background, size is relative and therefore carries too little information regarding environmental statistics. To overcome this shortcoming, future studies might employ nonnumerical/artificial symbols (other than Klingon letters) to represent numerosity while maintaining the size manipulation, possibly through well-placed reference

points on the screen, which should accentuate the perceived/ implied differences among the variable stimulus sizes. Additionally, supra-second probe durations were used in the present temporal bisection task (e.g., Rammsayer & Verner, 2016; Vicario, 2011), in order to allocate enough time for the simultaneous processing of multiple dimensional modalities, allowing for a more robust potential interaction. Therefore, a natural extension of the present findings would be an attempt to replicate these findings while employing subsecond probe durations, as well. This could prove to be a valuable endeavor, given that models such as ATOM (Walsh, 2003) predict that the demonstration of cross-dimensional interactions between time, space, and number is a result of shared representational codes that are responsible for the transformations necessary to coordinate quick and efficient motor responses, many of which take place in subsecond ranges. Additionally, a static frame around the presented stimuli could reduce the relativity of physical size and instead make size a better source of quantifiable information. Such a design could also mimic our Experiments 1 and 3 by utilizing different sizes (i.e., surface areas) of varying shapes while controlling for the number of edges (e.g., a square, a rectangle, and a parallelogram, with four vertices each) as the nonsymbolic to-be-timed stimuli. Finally, the average degree of shift in PSE was rather small in the present study (i.e., ~ 70 ms, corresponding to 4% of the average PSE). Larger ratios between the numerals used (e.g., 1 vs. 9) might be a potential way to achieve larger effects in future studies.

Author note This study was supported by a New Agendas for the Study of Time and TÜBA (Turkish Academy of Sciences) GEBİP 2015 grant to F.B. The data and materials for all experiments are available at <https://tinyurl.com/yxoq8cvc>.

References

- Alards-Tomalín, D., Leboe-McGowan, J. P., Shaw, J. D., & Leboe-McGowan, L. C. (2014). The effects of numerical magnitude, size, and color saturation on perceived interval duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 555–566. <https://doi.org/10.1037/a0035031>
- Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, *26*, 340–354.
- Allan, L. G. (1984). Contingent aftereffects in duration judgments. *Annals of the New York Academy of Sciences*, *423*, 116–130. <https://doi.org/10.1111/j.1749-6632.1984.tb23422.x>
- Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: First- and second-order principles of subjective time. *Annual Review of Psychology*, *65*, 743–771.
- Apaydın, N., Üstün, S., Kale, E. H., Çelikağ, İ., Özgüven, H. D., Baskak, B., & Çiçek, M. (2018). Neural mechanisms underlying time perception and reward anticipation. *Frontiers in Human Neuroscience*, *12*, 115. <https://doi.org/10.3389/fnhum.2018.00115>
- Balci, F., & Gallistel, C. R. (2006). Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion? *Psychonomic Bulletin & Review*, *13*, 636–642. <https://doi.org/10.3758/BF03193974>
- Balci, F., & Simen, P. (2016). A decision model of timing. *Current Opinion in Behavioral Sciences*, *8*, 94–101.
- Beckmann, J. S., & Young, M. E. (2009). Stimulus dynamics and temporal discrimination: Implications for pacemakers. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 525–537. <https://doi.org/10.1037/a0015891>
- Bellos, A. (2015). *The grapes of math: How life reflects numbers and numbers reflect life*. New York: Simon & Schuster.
- Berger, A., & Hill, T. P. (2011). Benford's law strikes back: No simple explanation in sight for mathematical gem. *Mathematical Intelligencer*, *33*, 85–91.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. <https://doi.org/10.1163/156856897X00357>
- Brigner, W. L. (1986). Effect of perceived brightness on perceived time. *Perceptual Motor Skills*, *63*, 427–430. <https://doi.org/10.2466/pms.1986.63.2.427>
- Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics*, *57*, 105–116. <https://doi.org/10.3758/BF03211853>
- Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*, 1118–1140. <https://doi.org/10.3758/BF03205526>
- Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B*, *364*, 1831–1840.
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*, 755–765. <https://doi.org/10.1038/nrn1764>
- Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*, 216–228. [10.1037/0097-0922.3.2.216](https://doi.org/10.1037/0097-0922.3.2.216)
- Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PLoS ONE*, *11*, e158725. <https://doi.org/10.1371/journal.pone.0158725>
- Conson, M., Cinque, F., Barbarulo, A. M., & Trojano, L. (2008). A common processing system for duration, order and spatial information: Evidence from a time estimation task. *Experimental Brain Research*, *187*, 267–274.
- Coull, J. T., Vidal, F., Nazarian, B., & Macar, F. (2004). Functional anatomy of the attentional modulation of time estimation. *Science*, *303*, 1506–1508.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.
- Dormal, V., & Pesenti, M. (2012). Processing magnitudes within the parietal cortex. *Horizons in Neuroscience Research*, *8*, 107–140.
- Dormal, V., Seron, X., & Pesenti, M. (2006). Numerosity–duration interference: A Stroop experiment. *Acta Psychologica*, *121*, 109–124.
- Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*, 131–136. <https://doi.org/10.1016/j.conb.2008.06.002>
- Fontes, R., Ribeiro, J., Gupta, D. S., Machado, D., Lopes-Júnior, F., Magalhães, F., ... Teixeira, S. (2016). Time perception mechanisms at central nervous system. *Neurology International*, *8*, 5939. <https://doi.org/10.4081/ni.2016.5939>
- Gallistel, C. R. (1990). *The organization of behavior*. Cambridge: MIT Press.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74. [https://doi.org/10.1016/0010-0277\(92\)90050-R](https://doi.org/10.1016/0010-0277(92)90050-R)

- Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4, 59–65. [https://doi.org/10.1016/S1364-6613\(99\)01424-2](https://doi.org/10.1016/S1364-6613(99)01424-2)
- Gibbon, J., & Church, R. M. (1990). Representation of time. *Cognition*, 37, 23–54. [https://doi.org/10.1016/0010-0277\(90\)90017-E](https://doi.org/10.1016/0010-0277(90)90017-E)
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52–77.
- Gil, S., & Droit-Volet, S. (2011). “Time flies in the presence of angry faces” . . . depending on the temporal task used! *Acta Psychologica*, 136, 354–362.
- Goodman, S. N. (1999). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, 130, 1005–1013.
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72, 561–582. <https://doi.org/10.3758/APP.72.3.561>
- Hayashi, M. J., Kanai, R., Tanabe, H. C., Yoshida, Y., Carlson, S., Walsh, V., & Sadato, N. (2013). Interaction of numerosity and time in prefrontal and parietal cortex. *Journal of Neuroscience*, 33, 883–893.
- Held, L., & Ott, M. (2018). On *p*-values and Bayes factors. *Annual Review of Statistics and Its Application*, 5, 393–419. <https://doi.org/10.1146/annurev-statistics-031017-100307>
- Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, 10, 389–395. <https://doi.org/10.3758/BF03202431>
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6, 435–448. <https://doi.org/10.1038/nrn1684>
- Hurewitz, F., Gelman, R., & Schneider, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, 103, 19599–19604.
- JASP Team (2018). JASP (Version 0.9)[Computer software].
- Jeffreys, H. (1961). *Theory of probability* (3rd). Oxford: Oxford University Press.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs: Prentice-Hall.
- Kallai, A. Y., & Tzelgov, J. (2012). The place-value of a digit in multi-digit numbers is processed automatically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1221–1233. <https://doi.org/10.1037/a0027635>
- Kaneko, S., & Murakami, I. (2009). Perceived duration of visual motion increases with speed. *Journal of Vision*, 9(7), 14. <https://doi.org/10.1167/9.7.14>
- Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron*, 53, 427–438.
- Karşilar, H., Kisa, Y. D., & Balci, F. (2018). Dilation and constriction of subjective time based on observed walking speed. *Frontiers in Psychology*, 9, 2565. <https://doi.org/10.3389/fpsyg.2018.02565>
- Kroger-Costa, A., Machado, A., & Santos, J. A. (2013). Effects of motion on time perception. *Behavioural Processes*, 95, 50–59.
- Levy, J. M., Namboodiri, V. M., & Hussain Shuler, M. G. (2015). Memory bias in the temporal bisection point. *Frontiers in integrative neuroscience*, 9, 44.
- Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences*, 25, 126–133. <https://doi.org/10.1016/j.lindif.2013.02.001>
- Martin, B., Wiener, M., & van Wassenhove, V. (2017). A Bayesian perspective on accumulation in the magnitude system. *Scientific Reports*, 7, 630. <https://doi.org/10.1038/s41598-017-00680-0>
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203–220. <https://doi.org/10.1037/h0087426>
- Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 303–313. <https://doi.org/10.1037/a0019961>
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320–334. <https://doi.org/10.1037/0097-7403.9.3.320>
- Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 591–597. <https://doi.org/10.1037/0097-7403.11.4.591>
- Meck W. H., Penney, T. B., & Pouthas, V. (2008). Cortico-striatal representation of time in animals and humans. *Current Opinion in Neurobiology*, 18, 145–152. <https://doi.org/10.1016/j.conb.2008.08.002>
- Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience*, 36, 313–336.
- Mo, S. S., & Michalski, V. A. (1972). Judgment of temporal duration of area as a function of stimulus configuration. *Psychonomic Science*, 27, 97–98.
- Montemayor, C., & Balci, F. (2007). Compositionality in language and arithmetic. *Journal of Theoretical and Philosophical Psychology*, 27, 53–72. <https://doi.org/10.1037/h0091281>
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215, 1519–1520. <https://doi.org/10.1038/2151519a0>
- Muller, T., & Nobre, A. C. (2014). Perceiving the passage of time: neural possibilities. *Annals of the New York Academy of Sciences*, 1326, 60–71.
- Okrand, M. (1992). *The Klingon dictionary*. New York: Simon & Schuster.
- Oliveri, M., Koch, G., & Caltagirone, C. (2009). Spatial-temporal interactions in the human brain. *Experimental Brain Research*, 195, 489–497. <https://doi.org/10.1007/s00221-009-1834-1>
- Oliveri, M., Vicario, C. M., Salerno, S., Koch, G., Turriziani, P., Mangano, R., . . . Caltagirone, C. (2008). Perceiving numbers alters time perception. *Neuroscience Letters*, 438, 308–311. <https://doi.org/10.1016/j.neulet.2008.04.051>
- Ono, F., & Kawahara, J.-I. (2007). The subjective size of visual stimuli affects the perceived duration of their presentation. *Perception & Psychophysics*, 69, 952–957. <https://doi.org/10.3758/BF03193932>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. <https://doi.org/10.1163/156856897X00366>
- Penney, T. B. (2003). Modality differences in interval timing: attention, clock speed, and memory. In W. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 209–234). Boca Raton: CRC Press.
- Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, 53, 293–305. <https://doi.org/10.1016/j.neuron.2006.11.022>
- Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, 41, 983–993.
- Rammesayer, T. H., & Verner, M. (2016). Evidence for different processes involved in the effects of nontemporal stimulus size and numerical digit value on duration judgments. *Journal of vision*, 16(7), 13-13.
- Simen, P., Rivest, F., Ludvig, E. A., Balci, F., & Killeen, P. (2013). Timescale invariance in the pacemaker-accumulator family of timing models. *Timing & Time Perception*, 1, 159–188.
- Thomas, E. A. C., & Cantor, N. E. (1976). Simultaneous time and size perception. *Perception & Psychophysics*, 19, 353–360.

- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock”. *Psychological Monographs: General and Applied*, 77(13, Whole No. 576), 1–31. [10.1037/h0093864](https://doi.org/10.1037/h0093864)
- Vicario, C. M. (2011). Perceiving numbers affects the subjective temporal midpoint. *Perception*, 40, 23–29.
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2018a). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... Morey, R. D. (2018b). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25, 35–57. <https://doi.org/10.3758/s13423-017-1343-3>
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7, 483–488. <https://doi.org/10.1016/j.tics.2003.09.002>
- Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. *The Quarterly Journal of Experimental Psychology*, 48(4), 289–310.
- Wearden, J. H. (1999). “Beyond the fields we know . . .”: Exploring and developing scalar timing theory. *Behavioural Processes*, 45, 3–21.
- Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, 61, 569–587. <https://doi.org/10.1080/17470210701282576>
- Xuan, B., Chen, X. C., He, S., & Zhang, D. R. (2009). Numerical magnitude modulates temporal comparison: An ERP study. *Brain Research*, 1269, 135–142.
- Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, 7(10), 2.1–5. <https://doi.org/10.1167/7.10.2>
- Zakay, D., & Block, R. A. (1995). An attentional gate model of prospective time estimation. In V. de Keyser, G. d’Ydewalle, & A. Vandierendonck (Eds.), *Time and the dynamic control of behavior* (pp. 167–178). Seattle: Hogrefe & Huber.
- Zakay, D., & Block, R. A. (1996). The role of attention in time estimation processes. In *Advances in psychology* (Vol. 115, pp. 143–164). Amsterdam, The Netherlands: North-Holland.
- Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science*, 6, 12–16.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.