CrossMark

Counting does not improve the accuracy of long time productions

Sven Thönes[1,2,3] · Heiko Hecht[2]

Abstract It is common to use counting strategies to produce time intervals accurately. Does counting improve the accuracy (deviation of produced duration from veridical duration) and precision (variability of produced duration) of time productions in the range of seconds to minutes? In a series of experiments, we compared chronometric counting to intuitive timing (no counting) and to attentional control (simultaneously performing mental arithmetic). In a field experiment, participants had to produce time intervals of 60-s duration in a classroom setting. Relative to intuitive timing, counting did not improve the accuracy (absolute error) of time productions but led to overproduction of duration (larger constant error). In four laboratory experiments, we tested the effects of counting on time production of 10-, 30-, 45-, 60-, and 90-s intervals. Consistently, counting did not improve the accuracy of time production. The relative overproduction due to counting was replicated for long intervals (60 and 90 s) but disappeared at shorter intervals. However, across all intervals tested, counting had positive effects on the precision of time production. As expected, mental arithmetic impaired accuracy and precision and led to overproduction of duration, indicating that participants followed instructions. Based on the experimental data, the overproduction of longer intervals due to counting can neither be explained in terms of attentional processes nor by means of a word-length effect when counting multisyllabic numbers or when participants switch their pace of counting from mono- to multisyllables.

Keywords Time perception · Time production · Counting · Accuracy · Precision

✉ Sven Thönes
thoenes@ifado.de

Heiko Hecht
hecht@uni-mainz.de

1 Leibniz Research Centre for Working Environment and Human Factors, Ardeystr. 67, 44139 Dortmund, Germany

2 Experimental Psychology, Department of Psychology, Johannes Gutenberg-Universität Mainz, Mainz, Germany

3 Occupational and Engineering Psychology, Department of Psychology, Technische Universität Darmstadt, Darmstadt, Germany

Introduction

In many everyday situations, we have to judge short durations accurately, such as when boiling eggs or brewing tea. Does counting help when we have to produce temporal intervals in this range? Most people spontaneously start to count when they have to produce duration. The experimental studies that have reported benefits of a counting strategy on temporal judgments have focused on rather short durations in the range of few seconds and used methods of *verbal estimation* (estimation of an elapsed time interval in chronometric units; Gilliland & Martin, 1940; Hicks & Allen, 1979a), *time reproduction* (reproduction of an elapsed time interval by timed motor responses; Getty, 1976; Grondin & Killeen, 2009b; Grondin, Laflamme, & Mioni, 2015; Hinton & Rao, 2004; Rattat & Droit-Volet, 2012), or *duration discrimination* (comparison of elapsed time intervals; Grondin, Meilleur-Wells, & Lachance, 1999; Grondin, Ouellet, & Roussel, 2004; Rattat & Droit-Volet, 2012; Wearden, Denovan, Fakhri, & Haworth, 1997) but never *time production* (indication of when a defined time interval has elapsed). We believe that a *time production* task and the use of longer intervals most

closely mimic the above-mentioned everyday situations. Therefore, we have investigated the potential effects of counting on *time production* of 10-, 30-, 45-, 60-, and 90-second intervals in a series of experiments while controlling for attentional processes.

Besides common habits of counting strategies, neuroclinical studies, for example in the field of developmental dyscalculia (Cappelletti, Freeman, & Cipolotti, 2011; Gilaie-Dotan, Rees, Butterworth, & Cappelletti, 2014), provide evidence for shared mechanisms between number and time processing, thus pointing to basic associations between counting and judgments of duration. Theoretical psychological models of interval timing, however, such as the pacemaker-accumulator model (Gibbon, Church, & Meck, 1984; Treisman, 1963), reveal the complexities involved when attempting to predict effects of chronometric counting on duration judgments. Basically, the pacemaker-accumulator model assumes an internal clock consisting of a pacemaker that emits pulses and an accumulator (or counter) collecting these pulses. As soon as a participant begins to process a time interval, an attentionally modulated switch between pacemaker and accumulator closes. Therefore, the clock pulses emitted by the pacemaker can reach the accumulator, which starts to collect the pulses. The more pulses are accumulated, the longer the perceived length of an interval. The amount of pulses collected in the accumulator is constantly compared to duration samples that are stored in reference (e.g., long term) memory. In a time production task, where a time interval is defined in terms of chronometric temporal units, e.g., "30 s," and the participant is required to produce the interval, for instance by giving two motor responses marking its beginning and end, the participant needs to refer to long-term memory representations of duration and repeatedly compare these with the perceived elapsed duration accumulated and stored in working memory. As soon as the amount of accumulated pulses matches the memory representation of the interval to be produced, the participant will decide to mark the end of the interval.

The accumulation of pulses requires attentional resources. If attention is distracted from the time perception task, for example due to a dual task, duration is underestimated and interval productions become longer and more variable (Brown, 1997, 2008). In terms of the internal clock model, attentional effects have been formalized, for example, by means of an attentional gate between pacemaker and accumulator (Block & Zakay, 1996; Zakay & Block, 1996). According to the attentional gate model, the pulses from the pacemaker need to pass the gate in order to reach the accumulator. The wider the gate opens, i.e., the more attention is directed to time, the more pulses can pass causing longer time estimates and shorter time productions. Note that the attentional gate is different from the switch, which needs to be (fully) closed to let any pulses reach the accumulator.

Why then should chronometric counting increase the accuracy (minimize the deviation of produced duration from the veridical duration) of duration judgments? According to Vierordt (1868), human participants overestimate short durations and underestimate long durations. This notion has been confirmed frequently, and empirical studies suggest the existence of an indifference interval somewhere between 0.4 and 5 s devoid of tendencies to over- or underestimate duration (Eisler, Eisler, & Hellström, 2008; Jones & McAuley, 2005; Woodrow, 1934). The exact location of the indifference interval is still controversial, and recent research suggests that it depends on the particular experimental context, for instance the temporal intervals presented to the participants (Grondin, 2010; Jones & McAuley, 2005). Irrespective of its exact location, according to the idea of an indifference interval somewhere around 1 s up to several seconds, judgments of duration in this range are more accurate than judgments of longer durations, for example 10-, 30-, or 60-s intervals, which are generally underestimated and overproduced. Subdividing such long intervals into a series of short (1-s) intervals that are closer to the indifference interval than the long duration to be judged therefore should increase the accuracy of the whole duration judgment. The sum of the constant errors of the produced subintervals is simply lower than the constant error of the undivided (whole) interval. At the same time, however, counting does not only involve the subdividing of a long interval into smaller pieces, but also the subsequent summing-up of these pieces. This task may require enough attentional resources to distract from the timing task itself. Especially in the case of long intervals, which involve the summing-up of many 1-s units, an attentional effect may disturb the accumulation process in the manner described above. An underestimation and overproduction of duration would be the result. In terms of the pacemaker-accumulator model, counting may reduce the error emerging from the pacemaker, while at the same time increasing the accumulation error, especially when the interval to be judged becomes longer, i.e., when the number of counts increases. Therefore, based on the pacemaker-accumulator model, we do not know whether chronometric counting improves or compromises the accuracy of duration judgments in the range of several seconds to 1 minute. The foregoing arguments most likely predict positive effects of counting at short time intervals of a few seconds, but impaired production of longer durations due to the attentionally demanding character of counting (by summing-up many 1-second units).

In previous studies on the effects of counting, rather short intervals in the range of a few seconds (or below) have typically been used. Several studies employing duration discrimination tasks have reported that counting improves temporal accuracy for stimuli lasting several seconds (Grondin et al., 1999; Grondin et al., 2004; Rattat & Droit-Volet, 2012; Wearden et al., 1997). Three studies focused on somewhat longer intervals up to 27 seconds providing mixed results: Whereas Gilliland and Martin (1940) and Hinton and Rao

(2004) reported no positive effects of counting on temporal accuracy in verbal time estimation, and time reproduction, respectively, data by Grondin and Killeen (2009b) indicate accuracy benefits when participants adopt counting strategies when reproducing intervals of 6 to 24 s (see also Grondin & Killeen, 2009a). To our knowledge, only one study has used durations approaching one minute (Hicks & Allen, 1979a; reported negative effects of counting on accuracy of verbal estimates), and no study has used the method of time production, which most closely mimics typical waiting periods.

In a first (field) experiment, we therefore investigated whether counting leads to more or less accurate *productions* of a 60-s interval compared with intuitive judgment without counting. In a third condition, we additionally asked the participants to engage in mental arithmetic while producing the time interval of 60 s. As pointed out earlier, such a cognitively demanding task distracts attention from the timing task, which is well-known to lead to temporal overproduction (Block, 1990; Zakay & Block, 1996). The purpose of this task was to ensure that participants followed instructions. Only when we replicate this effect can we interpret a potential null-effect of counting.

By means of four controlled experiments, we aimed at replicating the results from the first experiment in a lab situation and added four more target durations (10-, 30-, 45-, and 90-s intervals). Moreover, the experiments comprised several trial repetitions per condition, thus providing information not only about accuracy of time productions but also about their precision. While *accuracy* indexes the deviation of the produced time interval from the veridical value of the time interval, *precision* refers to the intra-individual *variability* of judgments across several trials. Studies on the effects of counting on precision of duration judgments indicate less variable (more precise) judgments when participants use counting strategies (Getty, 1976; Grondin & Killeen, 2009b; Grondin et al., 1999; Grondin et al., 2004; Hinton & Rao, 2004; Killeen & Weiss, 1987; Rattat & Droit-Volet, 2012; Wearden & Lejeune, 2008). These existing studies, however, have focused on durations in the range up to few seconds only. Similar to the accuracy of productions of longer time intervals, the positive effects of counting on temporal precision may disappear due to the attentional demands of counting. We therefore also investigated whether counting improves or compromises the precision of duration judgments in the range up to 90 s.

Experiment 1

Method

Participants

The sample for the study was drawn from an undergraduate student population (approximately 80% female, mean age 22

years). The experiment was conducted at the beginning of a psychology lecture. The lecture room contained 110 seats. The students were informed about the task procedure, and they were asked to decide whether or not they wanted to participate in the experiment. Two video cameras were used to record the simultaneous performance of 58 students who consented to participate.

Procedure

The participants' task was to produce time intervals of 60 s. In response to a start signal given by the experimenter, the participants were instructed to keep track of time and to silently raise their hand holding up a card when they thought that 60 s had elapsed. After the cards had been given to the participants, they were instructed to keep their eyes closed and to remain silent during the interval production task. Interval production was repeated twice resulting in three trials per participant in total. The participants successively performed three different *Tasks* during the interval productions. In the condition intuitive timing, they were instructed to produce the time interval of 60 s without counting or any other potential strategy. In the condition counting, the participants were instructed to count from 1 to 60 (one number per second) while producing the time interval. In the condition arithmetic (attentional control), the participants had to count back from 1,000 in steps of 7 while producing the time interval. The cards the participants had to raise in order to mark the end of each interval were either green, or yellow, or red. Based on the color of the cards, which were randomly distributed across the class (no clustering of groups), each participant was randomly assigned to one of three *Task order*-conditions. As realizing all possible task orders would have been too complex in the field setting, the experimental design followed a Latin square: a) participants who received a green card were instructed to count during the first interval, to time intuitively during the second, and to perform the arithmetic task during the third; b) participants holding a yellow card had to time intuitively first, followed by arithmetic and counting; c) a red card indicated the *Task order* arithmetic – counting – intuitive timing. Note that we did not expect any effect of *Task order* on produced duration. The purpose of manipulating *Task order* was to prevent a confounding of *Task* and trial. For example, if all participants had to count in the last trial (no manipulation of *Task order*), and if these productions were particularly accurate, the high level of accuracy could be explained by the fact that participants used the counting strategy as well as by the possibility that productions from the last trial may be most accurate due to learning.

Each trial ended after all participants had raised their hands. The experimenter then started the next trial, that is, he gave the start signal for the next 60-s interval to be produced.

Video analysis

Based on the video recordings, the produced durations in seconds were determined for each participant in each trial / *Task* (intuitive timing, counting, arithmetic). The response coder was blind to the hypotheses of the experiment. For each participant, an interval was defined as being produced when the hand raise was fully executed, that is, when the upward movement of the arm had stopped. Data from two participants were excluded from the analyses because in one trial their hands (cards) were not sufficiently visible.

Results

The data were analyzed in terms of the *constant error* (CE) and the *absolute error* (AE), which both refer to the accuracy of duration judgments. The CE is defined as the (signed) difference between the produced duration (PD) and the target duration (TD; 60 s): CE = PD – TD. The AE is determined by dividing the unsigned difference between PD and TD by TD: AE = |PD – TD| / TD (Mioni, Stablum, McClintock, & Grondin, 2014). Whereas the CE provides information about the direction of errors in duration judgments between different experimental conditions (positive values: relative overproduction; negative values: relative underproduction), the AE indexes the absolute (unsigned) discrepancy of duration judgments from the veridical duration. An AE of 0 indicates accurate duration judgments.

As a function of *Task* and *Task order*, mean CE and AE are presented in Fig. 1 (see also Supplementary Material Table 1).

The data were analyzed statistically by means of rmANOVAs, including the within-subjects factor *Task* and the between-subjects factor *Task order*. Huynh-Feldt-corrected values are reported. For subsequent pairwise comparisons, we adjusted the alpha-level according to the Bonferroni-Holm procedure (Holm, 1979), and report Cohen's d_z (Cohen, 1988) as a measure of effect size in a dependent measures design.

Constant error

There was a significant effect of *Task* on the CE, $F(1.75, 92.51) = 20.14$, $p < 0.001$, $\varepsilon = 0.87$, $_{partial}\eta^2 = 0.28$. As indicated by post-hoc t tests for dependent samples, the CE was significantly smaller for the condition intuitive timing relative to counting, $t(55) = 2.56$, $p = 0.013$, $d_z = 0.34$, as well as compared to arithmetic, $t(55) = 5.21$, $p < 0.001$, $d_z = 0.70$. Moreover, the CE were significantly smaller for counting as compared to arithmetic, $t(55) = 3.46$, $p = 0.001$, $d_z = 0.46$.

There was no main effect of *Task order*, $F(2, 53) = 0.35$, $p = 0.710$, $_{partial}\eta^2 = 0.01$. A significant interaction between *Task* and *Task order*, $F(3.49, 92.51) = 5.94$, $p = .001$, $\varepsilon = 0.87$, $_{partial}\eta^2 = 0.18$, indicated that the CE was particularly small ($M = 0.88$ s) and the effect of *Task* was most pronounced when intuitive timing was to be performed first. We decomposed the interaction by analyzing the effect of *Task* separately for each level of the factor *Task*
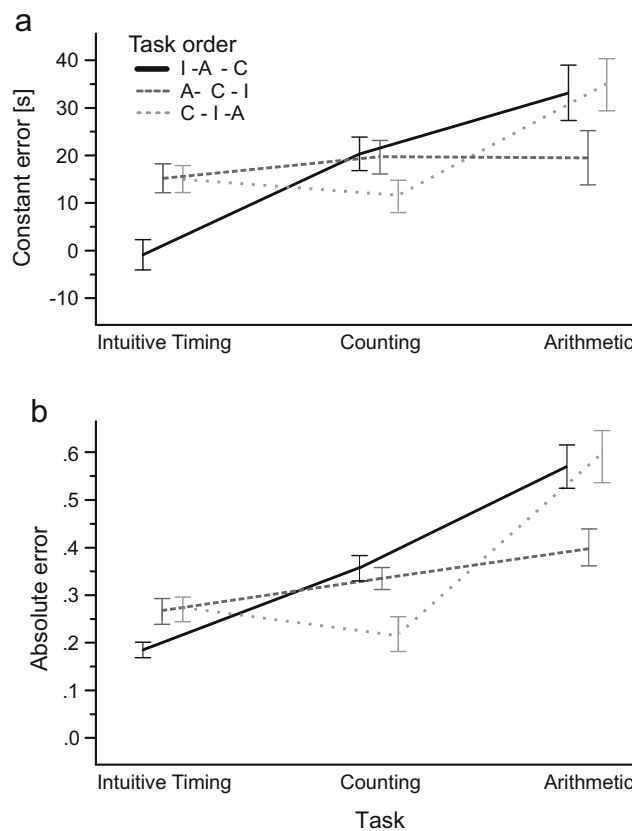
Fig. 1 CE in seconds (a) and AE (b) of produced duration as a function of *Task* and *Task order* in Experiment 1. Error bars indicate standard errors of the mean. Task order abbreviations I, C, and A indicate intuitive timing, counting, and mental arithmetic, respectively

order (three rmANOVAs). The effect of *Task* was significant for the task orders intuitive timing – arithmetic – counting, $F(1.72, 27.48) = 16.88$, $p < 0.001$, $\varepsilon = 0.86$, $_{partial}\eta^2 = 0.51$, and counting – intuitive timing – arithmetic, $F(1.81, 36.15) = 11.63$, $p < 0.001$, $\varepsilon = 0.90$, $_{partial}\eta^2 = 0.37$, but not for arithmetic – counting – intuitive timing, $F(1.68, 28.50) = 0.57$, $p = 0.54$, $\varepsilon = 0.83$, $_{partial}\eta^2 = 0.03$. Post-hoc paired samples t tests revealed significantly shorter productions for intuitive timing as compared to counting for the task order intuitive timing – arithmetic – counting, $t(16) = 5.04$, $p < 0.001$, $d = 1.22$, but not for counting – intuitive timing – arithmetic, $t(20) = 0.86$, $p = 0.400$, $d = 0.19$.

The CE appeared to generally increase from trial to trial. Thus, we included the within-subjects factor *Trial* in an additional rmANOVA analyzing whether time intervals were overproduced less in the first trial compared with later trials. This was indeed the case as witnessed by a main effect of *Trial*, $F(2, 110) = 8.75$, $p < 0.001$, $_{partial}\eta^2 = 0.14$. The productions in the first trial were shorter ($M = 10.36$ s, $SD = 19.66$ s) than those in the second ($M = 22.02$ s, $SD = 17.56$ s) and third trial ($M = 24.20$ s, $SD = 20.95$ s).

As these differences in the CE between first and later trials may modulate the effects of *Task*, we additionally analyzed task-dependent differences in the CE of the first trial only. A one-way ANOVA revealed a significant and large between-

subjects effect of *Task*, $F(2, 55) = 5.56$, $p = 0.006$, with shorter interval productions by intuitively timing participants compared with counting participants, $t(36) = 2.47$, $p = 0.019$, $d = 0.82$, and those who engaged in arithmetic, $t(33) = 3.18$, $p = 0.003$, $d = 1.12$ (unpaired samples t tests).

Absolute error

There was a significant effect of *Task* on the AE, $F(1.45, 86.38) = 17.55$, $p < 0.001$, $\varepsilon = 0.81$, $_{partial}\eta^2 = 0.25$. As indicated by post-hoc t tests for dependent samples, productions were not significantly more accurate for the condition intuitive timing relative to counting, $t(55) = 1.56$, $p = 0.124$, $d_z = 0.20$. Productions were significantly more accurate for intuitive timing compared with arithmetic, $t(55) = 4.87$, $p < 0.001$, $d_z = 0.65$, as well as for counting compared to arithmetic, $t(55) = 4.01$, $p < 0.001$, $d_z = 0.54$. There was no main effect of *Task order* on produced duration, $F(2, 53) = 0.22$, $p = 0.80$, $_{partial}\eta^2 = 0.01$. The interaction between *Task* and *Task order* did not reach statistical significance, $F(3.26, 86.38) = 2.33$, $p = 0.075$, $\varepsilon = 0.82$, $_{partial}\eta^2 = 0.08$.

As with the CE, we included the within-subjects factor *Trial* in an additional rmANOVA analyzing whether time intervals were produced more accurately in the first trial compared with later trials. This was confirmed by a significant main effect of *Trial*, $F(2, 110) = 3.66$, $p = 0.029$, $_{partial}\eta^2 = 0.06$. The productions in the first trial were closer on target ($M = 0.27$, $SD = 0.26$) than those in the second ($M = 0.38$, $SD = 0.28$) and third trial ($M = 0.41$, $SD = 0.34$).

Again, as these differences in accuracy between first and later trials may modulate the effects of *Task*, we additionally analyzed task-dependent differences in time productions of the first trial only. A one-way ANOVA revealed a significant between-subjects effect of *Task*, $F(2, 55) = 3.75$, $p = 0.030$. However, interval productions by participants who timed intuitively were not significantly more accurate than productions by those who counted, $t(36) = 0.54$, $p = 0.592$.

Discussion

We tested whether chronometric counting improves or impairs the accuracy of time production of a 60-s time interval, comparison with a no-counting condition (intuitive interval production), and a dual-task condition (mental arithmetic). In terms of the CE, the temporal judgments in the intuitive condition were approximately 10 seconds shorter than temporal judgments in the counting condition. This effect was most pronounced and largest in size when the participants were unbiased by previous interval productions and when intuitive timing was to be performed first (task order: intuitive timing – arithmetic – counting). Based on the AE, the difference between intuitive timing and counting remained statistically insignificant, indicating no benefit in accuracy due to counting.

These results are quite surprising as the majority of previous studies reported positive effects of counting on the accuracy of duration judgments (for recent reviews, see Rattat & Droit-Volet, 2012; Wearden & Lejeune, 2008). As expected, in terms of both dependent variables, time productions were largest and least accurate when the participants had engaged in the cognitively demanding arithmetic task, indicating that participants followed instructions (successful manipulation).

Interestingly, the produced intervals became larger by more than 10 seconds in the second and third trial compared with the first trial. A related repetition effect has been described previously (Hicks & Allen, 1979a, b; Ryan, 2011). This temporal overproduction (equivalent to underestimation in estimation tasks) in trials occurring later in the experiment may be explained by a decrease in the arousal level from the first to the later trials, which has caused a slower pacemaker rate of the internal clock (Gibbon, 1977; Gibbon et al., 1984; Treisman, 1963) resulting in fewer pulses accumulated.

Due to some methodological limitations, the results from Experiment 1 should be interpreted with caution. While the simultaneous testing of many participants in an everyday situation provides a high level of efficiency and ecological validity, it does not guarantee the high level of experimental control a laboratory experiment can provide. For example, even though the participants were instructed to remain silent and keep their eyes closed, some participants may have heard others raise their response cards, which may have prompted them to respond more quickly. Due to reasons of feasibility in the field, the three different experimental tasks were not presented repeatedly. The inclusion of several trial repetitions would have provided more reliable results. Another issue is related to the analysis of the responses based on the video tapes. Deciding when a hand raise was fully executed contains a certain level of uncertainty especially when a participant's arm was only partially visible on the video tapes. This factor may have caused some additional noise in the data. However, none of these limiting factors appeared to be systematic enough to invalidate the study, which we followed up with a series of controlled experiments in the laboratory.

Experiment 2

In Experiment 2, we aimed at replicating the results from the first experiment in a controlled lab situation and additionally added a second interval durations of 30 s. Moreover, we presented several trial repetitions per condition and investigated the effects of counting on the precision of time production.

Method

Sample

A total of 24 students (16 females; mean age = 23.5 years, SD = 6.9 years) participated in the experiment in return for partial course credit. According to the criterion proposed by Tukey (1977), no outliers were detected. All participants gave informed, written consent according to the Declaration of Helsinki. All participants had normal or corrected-to-normal vision and hearing. Based on the experimental design and the effect-size estimates obtained in Experiment 1, the sample size of 24 participants was recommended by G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Apparatus

Participants were tested individually while seated in a room with dimmed light. Using the software Python 2.7, all instructions and stimuli were presented by a computer equipped with a dual core E5700 3GHz processor and a NVIDIA Quadro FX1400 graphics card. The screen size (Nec MultiSync 90F) was 19" and the resolution was 1280 x 1024 pixels at a display rate of 89 Hz. The auditory stimuli were presented via headphones (Ultrasone HFI-780). All responses were given by using the spacebar of the computer keyboard.

Stimuli, procedure, design

The participants had to produce intervals of 30- and 60-s duration. While producing a time interval, the participant was instructed to either count from 1 to 30/60, to perform mental arithmetic, or to time the interval intuitively without any counting strategy (as in Experiment 1). At the beginning of the experiment, the participant received detailed written instructions for the upcoming tasks. The physical duration of the intervals, for example as indicated by the presentation of two tones with an inter-onset interval of 30 or 60s, was never revealed to the participant, and no feedback was given throughout the experiment.

Each trial began with a short written instruction that indicated which interval duration had to be produced (30 vs. 60 s) and which specific task had to be performed (intuitive timing vs. counting from 1 to 30/60 vs. mental arithmetic: counting backwards from 1,000 in steps of 7). To proceed, the participant was instructed to press the response button. Subsequently, for 1.5 s, a white fixation cross appeared in the center of the black background screen. The fixation cross was followed by a sinus tone (1,000 Hz, 50 ms) marking the beginning of the time interval to be produced by the participant. The participant was instructed to indicate the end of the time interval by pressing the response button and to keep the eyes closed during the production of the interval. Simultaneously to the button press, the tone was

presented again, this time to mark the end of the interval. Subsequently, the next trial began with the presentation of the trial-specific instruction (e.g., counting; 30 s). Each trial was presented four times resulting in 24 (2 *Intervals* * 3 *Tasks* * 4) trials per participant.

The trials were ordered randomly, separately for each participant. The whole experiment lasted approximately 30 minutes. The experimenter was blind to the hypotheses and to the results from Experiment 1.

Results

As in Experiment 1, we analyzed the data in terms of CE and AE. Additionally, as a measure of precision, we analyzed the coefficient of variation (CV). The CV is defined as the standard deviation of produced duration divided by the mean produced duration: $CV = SD(PD)/M(PD)$. As a function of *Task* and *Interval*, mean CE, AE, and CV are presented in Fig. 2 (see also Supplementary Material Table 2).

By means of rmANOVAs, we analyzed the effects of *Task* and *Interval* statistically. Huynh-Feldt-corrected values are reported. For pairwise comparisons, we additionally report d_z as a measure of effect size in a dependent measures design (Cohen, 1988). For pairwise post-hoc comparisons between the different conditions of *Task*, we adjusted the α-level according to the Bonferroni-Holm procedure (Holm, 1979).

Constant error

There was a significant effect of *Task* on the CE, $F(2, 46) = 8.30$, $p = 0.001$, $_{partial}\eta^2 = 0.27$, indicating overproduction of time intervals in the arithmetic condition ($M = 11.24$ s, $SD = 10.72$ s) compared with counting ($M = 6.45$ s, $SD = 8.49$ s), and overproduction in counting compared to intuitive timing ($M = 2.80$ s, $SD = 12.49$ s). The effect of *Interval* did not reach statistical significance, $F(1, 23) = 3.40$, $p = 0.078$, $_{partial}\eta^2 = 0.13$. A significant interaction between *Task* and *Interval*, $F(2, 46) = 4.59$, $p = 0.015$, $_{partial}\eta^2 = 0.17$, suggested that differences between counting and intuitive timing are specific to the 60-s target interval. We decomposed the interaction by analyzing the effect of *Task* separately for both intervals (two additional rmANOVAs) and further analyzed the effect of *Task* by means of specific paired samples t tests. The effect of *Task* was significant for both intervals, 30 s: $F(2, 46) = 8.84$, $p = 0.001$, $_{partial}\eta^2 = 0.28$, 60 s: $F(2, 46) = 7.06$, $p = 0.002$, $_{partial}\eta^2 = 0.24$. The pairwise comparisons confirmed that counting caused a significant overproduction of the 60-s target interval, compared to intuitive timing, $t(23) = 2.68$, $p = 0.013$, $d_z = 0.55$. There was no such difference between counting and intuitive timing for the 30-s target duration, $t(23) = 0.26$, $p = 0.795$, $d_z = 0.05$. In comparison to counting, the 30-s target interval was significantly overproduced when participants performed mental arithmetic, $t(23) = 3.45$, $p = 0.002$, $d_z = $
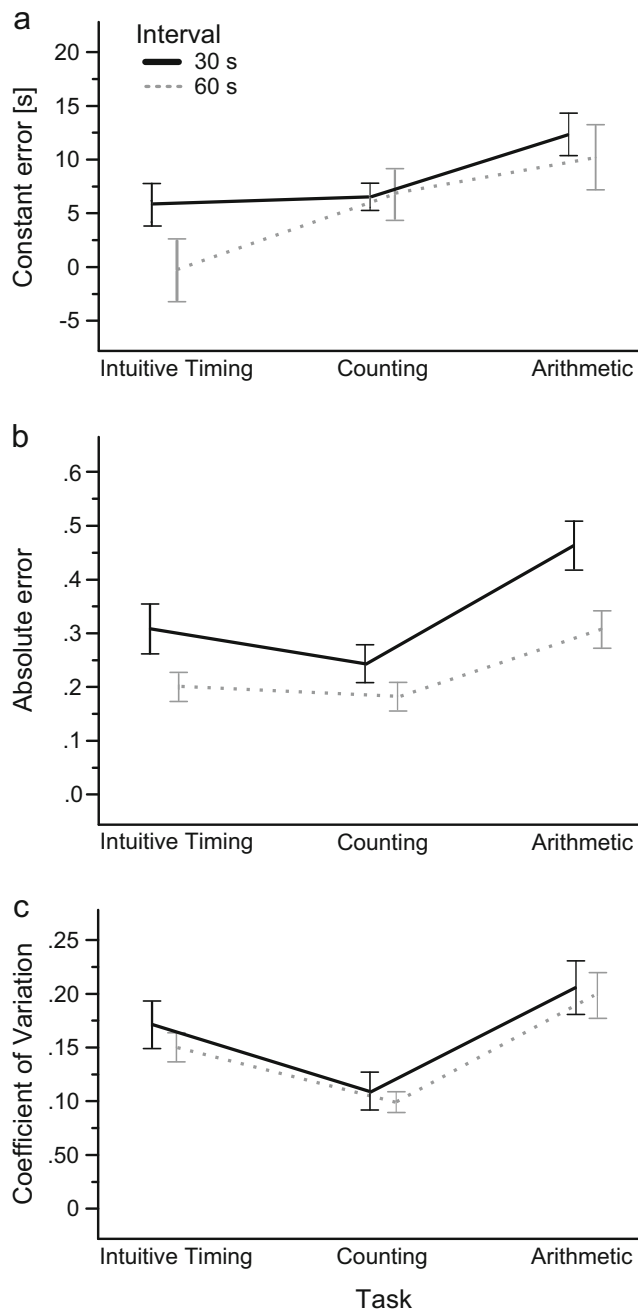
Fig. 2 CE in seconds (a), AE (b), and CV (c) of produced duration as a function of *Task* and *Interval* in Experiment 2. Error bars indicate standard errors of the mean

0.70. There was no significant difference between counting and arithmetic for the 60-s interval, $t(23) = 1.21$, $p = 0.238$, $d_z = 0.25$.

Absolute error

There was a significant effect of *Task* on the AE, $F(2, 46) = 9.54$, $p < 0.001$, $_{partial}\eta^2 = 0.29$, indicating more accurate interval production in the counting ($M = 0.21$, $SD = 0.15$) and

intuitive ($M = 0.26$, $SD = 0.14$) conditions compared with mental arithmetic ($M = 0.38$, $SD = 0.21$). Regarding *Interval*, $F(1, 23) = 19.66$, $p < 0.001$, $_{partial}\eta^2 = 0.46$, produced duration was more accurate for the 60-s target interval ($M = 0.23$, $SD = 0.09$) compared with the 30-s target interval ($M = 0.34$, $SD = 0.17$). The interaction between *Task* and *Interval* remained nonsignificant, $F(2, 46) = 2.93$, $p = 0.063$, $_{partial}\eta^2 = 0.11$. We further analyzed the effect of *Task* by means of paired samples t tests. The pairwise comparisons showed no significant differences between counting and intuitive timing, 30 s: $t(23) = 1.30$, $p = 0.206$, $d_z = 0.27$, 60 s: $t(23) = 0.69$, $p = 0.500$, $d_z = 0.14$. Compared with counting (and intuitive timing), both target intervals were produced less accurately when participants performed mental arithmetic, 30 s: $t(23) = 4.22$, $p < 0.001$, $d_z = 0.86$, 60 s: $t(23) = 3.24$, $p = 0.004$, $d_z = 0.66$ (30 s: $t(23) = 3.21$, $p = 0.004$, $d_z = 0.65$, 60 s: $t(23) = 2.36$, $p = 0.027$, $d_z = 0.48$).

Coefficient of variation

In a rmAVOVA including the factors *Task* and *Interval*, there was a significant effect of *Task*, $F(2, 46) = 15.73$, $p < 0.001$, $_{partial}\eta^2 = 0.41$, indicating less precise duration production in the intuitive timing and arithmetic conditions as compared to counting. There was no significant effect of *Interval*, $F(1, 23) = 1.57$, $p = 0.223$, $_{partial}\eta^2 = 0.05$, nor an interaction between *Task* and *Interval*, $F(2, 46) = 0.03$, $p = 0.972$, $_{partial}\eta^2 < 0.01$. We further analyzed the effect of *Task* by means of paired samples t tests. The pairwise comparisons confirmed that mental arithmetic impaired the precision of duration production of the 30-s target interval, $t(23) = 3.37$, $p = 0.003$, $d_z = 0.69$, and the 60-s target interval, $t(23) = 3.71$, $p = 0.001$, $d_z = 0.76$, in comparison to counting. Descriptive differences in the CV between counting and intuitive timing did not reach statistical significance for the 30-s interval, 30 s: $t(23) = 2.07$, $p = 0.050$, $d_z = 0.42$. For the 60-s interval, time productions were significantly more precise in the counting condition, $t(23) = 2.77$, $p = 0.011$, $d_z = 0.57$.

Discussion

In terms of the CE, we replicated the result from Experiment 1. In comparison to intuitive timing, counting (similar to mental arithmetic) led to overproduction of time intervals. This result was carried by the long interval, as there were no differences in mean productions between intuitive timing and counting for the 30-s interval. Based on the AE, the accuracy of interval production did not differ between intuitive timing and chronometric counting. Across target durations, the precision of time production (CV) was higher when participants were instructed to count. The participants' productions were least precise when performing mental arithmetic. This result

was expected and confirms that our participants followed instructions.

Why did our participants consistently overproduce the 60-s time interval when applying a counting strategy? This may be because the mental production of larger numbers, for example subvocalizing "forty-seven," takes more time than the production of smaller numbers, such as "four" (Ellis, 1992). Such a word-length effect (Baddeley, Thomson, & Buchanan, 1975) may lead to slower counting in the range above 10 seconds, resulting in temporal overproduction. As mentioned in the *Introduction*, an alternative explanation would be that counting represents a (light) dual-task condition, especially when it comes to longer intervals where many 1-s subintervals need to be summed up.

Experiment 3

By instructing participants to count to 10 in different numerical ranges (1-10, 21-30, and 51-60), we tested whether the word length can account for the temporal overproduction in the counting condition. If word length was the relevant factor, overproduction of duration should be most pronounced in the condition counting from 51 to 60, because this range contains the longest words to be vocalized. Moreover, we compared the accuracy and precision of duration production of 10-s intervals between intuitive timing and (normal) counting from 1 to 10.

Method

The sample and the laboratory settings (apparatus) were identical to Experiment 2. The participant was instructed to produce time intervals of 10-s duration, either by counting from 1 to 10, from 21 to 30, or from 51 to 60, or to time the interval intuitively.

As in Experiment 2, each trail began with the presentation of a short instruction that indicated which specific task had to be performed while producing the 10-s time interval (intuitive timing vs. counting from 1 to 10 vs. counting from 21 to 30 vs. counting from 51 to 60). The participant had to press the response button to proceed after having read the instructions. As in Experiment 2, a fixation cross appeared for 1.5 s and was followed by the tone that marked the beginning of the interval. Again, the participant was instructed to indicate the end of the interval by pressing the response button and to keep eyes closed until the response. Simultaneously to the button press, the tone was presented again to mark the end of the interval. The next trial began with the presentation of the trial-specific instruction. Each trial was presented four times resulting in 16 (4 Tasks * 4) trials per participant.

The trials were ordered randomly, separately for each participant. The whole experiment lasted approximately 10
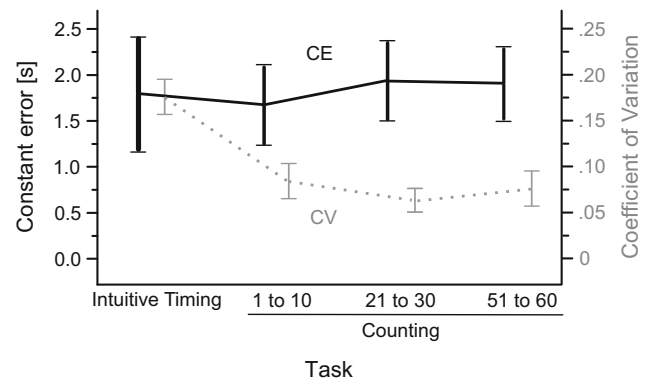

Fig. 3 CE (in seconds; solid line) and CV (dotted line) of produced duration as a function of *Task* in Experiment 3. Error bars indicate standard errors of the mean

minutes. The experimenter was blind to the hypotheses and to the results from Experiments 1 and 2.

Results

We analyzed the data in terms of CE and CV. As a function of *Task*, the CE in seconds and the CV of produced duration are presented in Fig. 3.

In an rmANOVA, we tested a possible effect of word length on the CE. Therefore, the factor *Task* included the three counting conditions (counting from 1 to 10 vs. counting from 21 to 30 vs. counting from 51 to 60). Huynh-Feldt-corrected values are reported. There was no significant effect of *Task*, $F(1.65, 37.89) = 1.09$, $p = 0.335$, $\varepsilon = 0.82$, $_{partial}\eta^2 = 0.05$, indicating no effect of word length on produced duration.

In a second step, we investigated whether the CE of the 10-s interval differed between normal counting (from 1 to 10) and intuitive timing. A paired-samples t test did not indicate such an effect, $t(23) = 0.24$, $p = 0.816$, $d_z = 0.05$.

In an additional t test, we compared the CV of produced 10-s duration (precision) between intuitive timing and normal counting. Interval productions were significantly more precise for counting compared with intuitive timing, $t(23) = 3.95$, $p = 0.001$, $d_z = 0.81$.

Discussion

To test whether the temporal overproduction in the long interval-counting conditions in Experiments 1 and 2 was caused by an effect of word length, we instructed our participants to count to 10 in three different ways: from 1 to 10, from 21 to 30, and from 51 to 60. If word length was the relevant factor, overproduction of duration was expected to be strongest in the counting from 51 to 60 condition, because this condition contains the longest words to be vocalized. However, there was clearly no effect of counting strategy on the CE. Based on this result, the overproduction of longer intervals cannot be explained in terms of a word-length effect.

As a side note, the lack of an effect questions the (German) habit to count from 21 to 30 to produce more accurately durations in the range below 10 s. It equally questions the (American) habit to append "Mississippi" to the count of 1, 2, 3, etc.

Experiment 4

In Experiment 4, we investigated whether the differences between intuitive timing and counting at 60-s intervals extend to interval durations of 45 and 90 s.

Method

Sample

A total of 25 students participated in the experiment in return for partial course credit. According to the criterion proposed by Tukey (1977), two far outliers were detected (participant 8: large CEs; participant 20: large CVs) and excluded from the analysis. The remaining sample consisted of 23 students (19 females; mean age = 22.8 years, SD = 4.8 years). All participants gave informed, written consent according to the Declaration of Helsinki. All participants had normal or corrected-to-normal vision and hearing.

Apparatus, procedure, design

Besides changing the durations to be produced from 30 and 60 s to 45 and 90 s, apparatus, experimental procedure, and design were identical to Experiment 2.

Results

The data analysis followed the same procedure as in Experiment 2. As a function of $Task$ and $Interval$, mean CE, AE, and CV are presented in Fig. 4 and Supplementary Material Table 3.

Constant error

There was a significant effect of $Task$ on the CE, $F(2, 44)$ = 4.14, p = 0.023, $_{partial}\eta^2$ = 0.16, indicating overproduction of time intervals in the arithmetic condition (M = 9.52 s, SD = 20.37 s) compared with counting (M = 6.85 s, SD = 13.24 s) and overproduction in counting compared to intuitive timing (M = –0.50 s, SD = 17.97 s). The effect of $Interval$ was also statistically significant, $F(1, 22)$ = 11.69, p = 0.002, $_{partial}\eta^2$ = 0.35, with a larger mean CE in the 45-s duration condition (M = 9.45 s, SD = 13.81 s) as compared to the 90-s duration condition (M = 1.13 s, SD = 20.57 s). A significant interaction between $Task$ and $Interval$, $F(2, 44)$ = 6.43, p = 0.004, $_{partial}\eta^2$
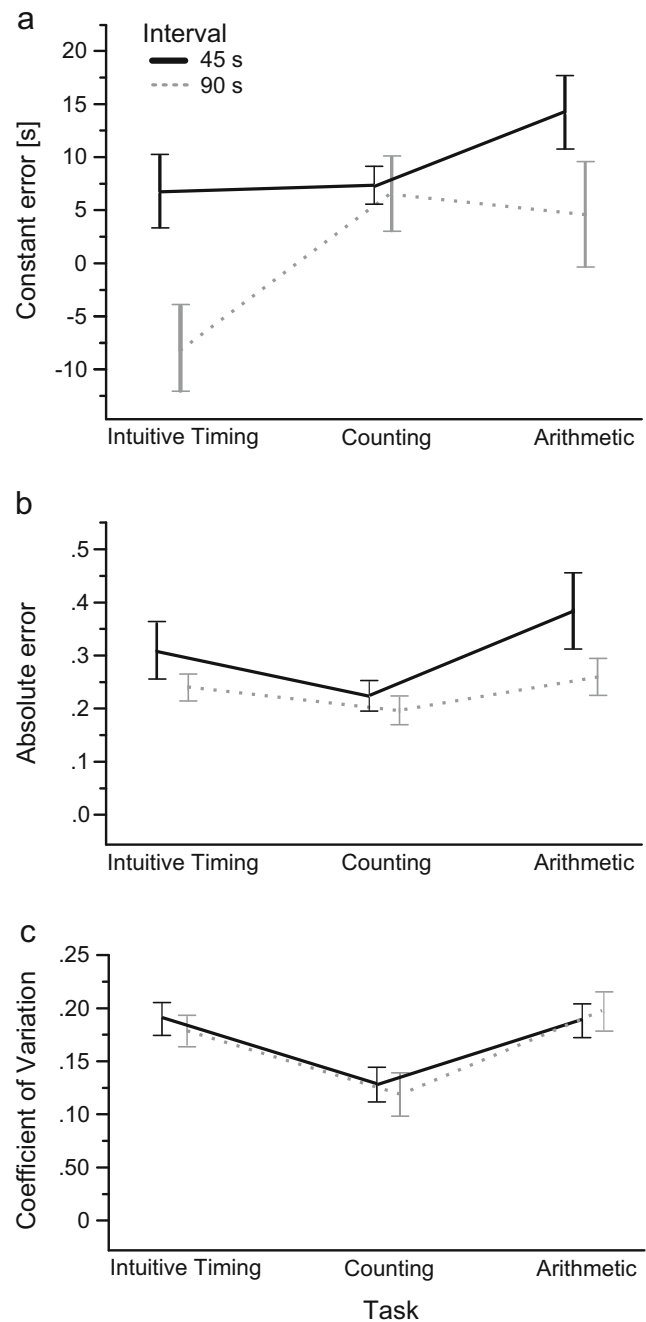
Fig. 4 CE in seconds (a), AE (b), and CV (c) of produced duration as a function of $Task$ and $Interval$ in Experiment 4. Error bars indicate standard errors of the mean

= 0.23, indicated that differences between counting and intuitive timing are carried by the 90-s target interval. We decomposed the interaction by analyzing the effect of $Task$ separately for both intervals (two additional rmANOVAs) and further analyzed the effect of $Task$ by means of paired samples t tests. The effect of $Task$ was significant for both intervals (in both additional rmANOVAs), 45 s: $F(2, 44)$ = 3.45, p = 0.041, $_{partial}\eta^2$ = 0.14, 90 s: $F(2, 44)$ = 5.23, p = 0.009, $_{partial}\eta^2$ = 0.19. The pairwise comparisons confirmed

that counting caused a significant overproduction of the 90-s target interval, compared to intuitive timing, $t(22) = 3.40$, $p = 0.003$, $d_z = 0.71$. There was no such difference between counting and intuitive timing for the 45-s target duration, $t(22) = 0.10$, $p = 0.923$, $d_z = 0.02$. In comparison to counting, the 45-s target interval was significantly overproduced when participants performed mental arithmetic, $t(22) = 2.12$, $p = 0.046$, $d_z = 0.44$. There was no significant difference between counting and arithmetic for the 90-s interval, $t(22) = 0.41$, $p = 0.688$, $d_z = 0.08$.

Absolute error

Task had a significant effect on AE, $F(2, 44) = 5.17$, $p = 0.010$, $_{partial}\eta^2 = 0.19$, indicating more accurate interval production in the counting ($M = 0.20$, $SD = 0.13$) and intuitive timing ($M = 0.27$, $SD = 0.18$) conditions compared with mental arithmetic ($M = 0.32$, $SD = 0.24$). There also was a significant effect of *Interval*, $F(1, 22) = 5.60$, $p = 0.027$, $_{partial}\eta^2 = 0.20$. Produced duration was more accurate for the 90-s target interval ($M = 0.23$, $SD = 0.13$) compared with the 45-s target interval ($M = 0.30$, $SD = 0.24$). The interaction between *Task* and *Interval* was not significant, $F(2, 44) = 1.23$, $p = 0.302$, $_{partial}\eta^2 = 0.05$. We further analyzed the effect of *Task* by means of paired samples t tests. The pairwise comparisons showed no significant differences between counting and intuitive timing, 45 s: $t(22) = 1.80$, $p = 0.085$, $d_z = 0.38$, 90 s: $t(22) = 1.49$, $p = 0.150$, $d_z = 0.31$. In comparison to counting, both target intervals were produced less accurately when participants performed mental arithmetic, 45 s: $t(22) = 2.57$, $p = 0.018$, $d_z = 0.54$, 90 s: $t(22) = 2.73$, $p = 0.012$, $d_z = 0.57$. There were no significant differences between mental arithmetic and intuitive timing, 45 s: $t(22) = 1.24$, $p = 0.229$, $d_z = 0.26$, 90 s: $t(22) = 0.74$, $p = 0.467$, $d_z = 0.15$.

Coefficient of variation

There was a significant effect of *Task* on CV, $F(1.90, 41.10) = 7.73$, $\varepsilon = 0.93$, $p = 0.002$, $_{partial}\eta^2 = 0.26$, indicating less precise duration production in the intuitive timing ($M = 0.18$, $SD = 0.09$) and arithmetic ($M = 0.19$, $SD = 0.10$) conditions compared with counting ($M = 0.12$, $SD = 0.09$). There was no effect of *Interval*, $F(1, 22) = 0.03$, $p = 0.871$, $_{partial}\eta^2 < 0.01$, and no interaction between *Task* and *Interval*, $F(2, 44) = 0.27$, $p = 0.768$, $_{partial}\eta^2 = 0.01$. We further analyzed the effect of *Task* by means of paired samples t tests. For both time intervals, the pairwise comparisons confirmed that counting enhanced the precision of duration production compared with intuitive timing, 45-s target interval: $t(22) = 2.56$, $p = 0.018$, $d_z = 0.40$, 90-s target interval: $t(22) = 3.43$, $p = 0.002$, $d_z = 0.71$. There were no differences in the CV between intuitive timing and mental arithmetic (t values < 1).

Discussion

The results from Experiment 4 are highly compatible with those obtained in Experiment 2 (and in Experiment 1). Compared with intuitive timing and similar to mental arithmetic, chronometric counting caused an overproduction of the long target duration (90 s). Again, this difference between intuitive timing and counting was not observable for the shorter interval (45 s). The accuracy of interval production, in terms of the AE, did not differ between intuitive timing and chronometric counting. And again, across both target durations, the precision of time production (CV) was higher when participants were instructed to count. The participants' productions were least precise when performing mental arithmetic, again confirming that our participants followed instructions.

Experiment 5

The overproduction of longer intervals due to counting was replicated in Experiment 4. Although Experiment 3 did not provide evidence for the assumed word-length effect, an involvement of word length cannot yet entirely be ruled out. The potential word-length effect may not be a universal effect but rather occur when the participant switches to counting multisyllables after having adopted a pace of counting that fits the use of monosyllables at the beginning of the count.[1] We tested this hypothesis by instructing our participants to count from 1 to 10 in three different ways: normal counting from 1 to 10, counting "1-2-3-4-5-6-27-28-29-10," and counting "1-2-3-24-25-26-27-28-29-10."

If switching to multisyllables was the relevant factor, the overproduction of duration should be pronounced in the two latter conditions, because these require the switching from monosyllables to words comprised of 4 and 5 syllables, respectively. Moreover, the overproduction should be strongest in the "1-2-3-4-5-6-27-28-29-10" condition, because before the switch, it induces more adaptation to a monosyllabic pace (until "6"; late switch) than the "1-2-3-24-25-26-27-28-29-10" condition (until "3"; early switch).

As in Experiment 3, we additionally compared the CEs and CVs of duration production of 10-s intervals between intuitive timing and (normal) counting from 1 to 10.

Method

The sample and the laboratory settings (apparatus) were identical to Experiment 3. The participant was instructed to produce time intervals of 10-s duration. During the interval production, the participant was instructed to count normally from

---

[1] We are grateful to the reviewer who pointed out this possibility.

1 to 10, to count 1-2-3-4-5-6-27-28-29-10, to count 1-2-3-24-25-26-27-28-29-10, or to produce the interval intuitively.

As in the previous experiments, each trial began with the presentation of a short instruction that indicated which specific task had to be performed while producing the 10-s time interval (intuitive timing vs. counting from 1 to 10 vs. counting 1-2-3-4-5-6-27-28-29-10, which will be referred to as counting 1 [27] 10, vs. counting 1-2-3-24-25-26-27-28-29-10, which will be referred to as counting 1 [24] 10). The participant had to press the response button to proceed after having read the instructions. A fixation cross appeared for 1.5 s and was followed by the tone that marked the beginning of the interval. Again, the participant was instructed to indicate the end of the interval by pressing the response button and to keep the eyes closed until response. Simultaneously with the button press, the tone was presented again to mark the end of the interval. The next trial began with the presentation of the trial-specific instruction. Each trial was presented four times resulting in 16 (4 Tasks * 4) trials ordered randomly per participant. The whole experiment lasted approximately 10 minutes. The experimenter was blind to the hypotheses and to the results from the previous experiments.

Results

As in Experiment 3, we analyzed the data in terms of CE and CV. As a function of *Task*, the CE in seconds and the CV of produced duration are presented in Fig. 5.

In an rmANOVA, we tested a possible effect of word length on the CE. Therefore, the factor *Task* included the three counting conditions (counting from 1 to 10 vs. counting 1 [27] 10 vs. counting 1 [24] 10). Huynh-Feldt-corrected values are reported. There was no significant effect of *Task*, $F(1.94, 42.65) = 1.25$, $p = 0.296$, $\varepsilon = 0.97$, $_{partial}\eta^2 = 0.05$, indicating no effect of pace switching on produced duration.

As in Experiment 3, in an additional step, we investigated whether the CE and the CV of the produced 10-s
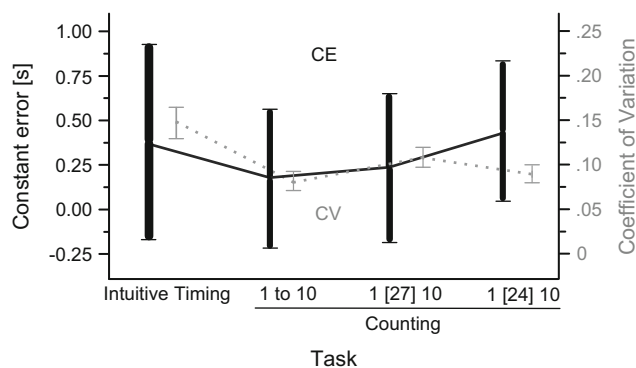

Fig. 5 CE (in seconds; solid line) and CV (dotted line) of produced duration as a function of *Task* in Experiment 5. Error bars indicate standard errors of the mean

intervals differed between normal counting (from 1 to 10) and intuitive timing. Paired-samples *t* tests did not indicate an effect on the CE, $t(22) = 0.70$, $p = 0.493$, $d_z = 0.15$, but once more interval productions were significantly more precise (CV) for counting as compared to intuitive timing, $t(22) = 4.52$, $p < 0.001$, $d_z = 0.94$.

Discussion

We tested whether the temporal overproduction in the long interval-counting condition (relative to intuitive timing) was caused by switching to multisyllables after having adopted a pace that fits the counting of monosyllables. The CE did not differ between normal counting and counting that required switching from monosyllables to words with 4 and 5 syllables. Moreover, there was no effect of degree of adaptation to monosyllabic counting (early vs. late switch) on produced duration. Based on these results, the overproduction of longer intervals cannot be explained in terms of a pace-switching effect.

In line with the results from Experiment 3, compared with intuitive timing, normal counting did not lead to an overproduction of the 10-s intervals but again facilitated the precision of time production.

General Discussion

In five experiments, we tested the effects of chronometric counting on the accuracy (and precision) of duration judgments in the range between 10 and 90 seconds, in comparison to a no-counting condition (intuitive interval production), and a dual-task condition (mental arithmetic). Beside practical arguments (the common use of counting strategies) indicating positive effects of counting on the accuracy of time production, theoretical considerations do not clearly predict whether counting should improve time production, especially of longer intervals. On the one hand, subdividing a long interval into several 1-s units should decrease the overall constant error, thus improving the accuracy of time production. On the other hand, the integrating aspect of counting longer durations (summing-up many single 1-s units) represents a cognitively demanding dual task that may disturb the accumulation process, which is necessary for the production of time intervals.

At 10-, 30-, and 45-s time intervals, mean produced durations (constant errors) did not differ between counting and intuitive timing. For the longer durations of 60 and 90 s, however, the constant errors indicated significant overproductions of duration of approximately 10 seconds due to counting compared with intuitive interval production. As expected, across experiments, time productions were largest when the participants engaged in the cognitively demanding arithmetic task. This result is compatible with the well-established and robust

effects of dual tasks on time perception (Brown, 1997; Champagne & Fortin, 2008; Rammsayer & Ulrich, 2011), thus indicating that participants had carefully followed the instructions in our experiments.

As an (unsigned) index of the absolute discrepancy of productions from the target duration (accuracy), the absolute error did not differ between counting and intuitive timing, whereas it was increased when participants performed the arithmetic task. These results were highly consistent across all our experiments and interval durations tested. Accordingly, counting did not improve the accuracy of time production across a wide range of long durations between 10 and 90 s.

In contrast to common belief and in contrast to the positive effects of counting on duration reproduction (Getty, 1976; Hinton & Rao, 2004) and discrimination (Grondin et al., 1999; Grondin et al., 2004; Wearden et al., 1997), our results do not indicate positive effects of counting on the accuracy of time production of longer intervals. They are, however, consistent with findings by Hicks and Allen (1979a) who reported underestimation of duration in a verbal estimation task when participants used a counting strategy. In comparison to the study by Grondin and Killeen (2009b), who let participants reproduce intervals between 6 and 24 s, we obtained quite similar constant errors in the intuitive timing conditions at 10 and 30 s. However, the results from the counting conditions are rather incompatible. This may be due to the different temporal tasks used (time production vs. time reproduction) and the special counting instructions in Grondin and Killeen (2009b), where participants had to adopt a 1 to 10 strategy at their subjectively preferred pace of counting. This probably resulted in constant errors being close to zero on average.

Why did our participants systematically overproduce the long durations when applying a chronometric counting strategy in comparison to intuitive timing? We had assumed that this may be the result of a word-length effect in the mental production of larger numbers. For example, subvocalizing "fif-ty-one" takes more time than the production of smaller numbers, such as "one" (Ellis, 1992). Such a word-length effect (Baddeley et al., 1975) may lead to slower counting in the range above 10 seconds resulting in temporal overproduction. In the third experiment, we explicitly tested this hypothesis by instructing participants to apply three different counting strategies, which were associated with the mental production of numbers of different length. We did not find evidence for an effect of word length on produced duration and therefore reject the word-length hypothesis. An alternative explanation for the overproduction of long intervals due to counting was related to a potential effect of switching from monosyllables to multisyllables on the pace of counting, and thus timing. In Experiment 5, we tested the pace-switching hypothesis by comparing normal counting from 1 to 10 with two conditions that included strong switches from monosyllables (e.g., "sechs" [six]) to words comprised of 4 and 5

syllables (e.g., "sie-ben-und-zwan-zig" [twen-ty-seven]). We did not find any differences in the CEs between normal counting and counting that included strong switches. Accordingly, we also rejected the pace-switching hypothesis as a potential explanation for the overproduction due to counting. In the three experiments testing longer durations, mean interval productions of 60 and 90 seconds were consistently comparable between the counting and the arithmetic condition (both conditions led to overproduction of duration). Accordingly, and in line with the assumption that the summing up of many 1-s units requires a substantial amount of attentional resources, counting in the range of larger numbers could be viewed as a dual-task condition that distracts attention from the timing task, thus causing temporal overproduction. This explanation, however, is challenged by the result that the variability as well as the absolute error of time production were lower in the counting condition as compared to mental arithmetic. A distraction of attention in the counting condition would be expected to go along with an increase in variability comparable to the mental arithmetic condition (for effects of attentional distraction on the precision of duration judgments, see, for example, Grondin, Laflamme, & Gontier, 2014). Moreover, if the summing-up of many 1-s units during counting was attentionally demanding, the constant errors in the counting conditions should be larger for longer (60- and 90-s) intervals compared with shorter (30- and 45-s) intervals. However, this was clearly not the case in our experiments. Accordingly, the overproduction of long intervals due to counting remains unexplained so far.

Across different interval durations, the precision of time production was clearly enhanced when participants were instructed to count. Interval productions were more variable when the participants produced the interval intuitively. These results are consistent with and extend the previous reports about positive effects of counting on the precision of short duration judgments that were limited to duration discrimination and reproduction tasks (Getty, 1976; Grondin et al., 1999; Grondin et al., 2004; Hinton & Rao, 2004; Rattat & Droit-Volet, 2012; Wearden & Lejeune, 2008). The mechanisms underlying the diverging effects of chronometric counting on the accuracy and precision of duration judgments need to be addressed by future research.

Taken together, chronometric counting has differential effects on the accuracy and precision of time production, and, partially, these effects depend on the interval durations that are to be produced. Whereas counting does not improve the accuracy of time productions across intervals of 10 to 90 s, the precision of duration judgments in this range is enhanced by counting. Based on the current data, the relative overproduction of long durations due to counting compared with intuitive timing can neither be explained in terms of word-length or pace-switching effects nor by an effect of attentional distraction. The consistent result that intuitive duration judgments are

surprisingly accurate supports the notion of an internal clock that does not require higher cognitive processes to judge duration accurately.

References

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575–589.

Block, R. A. (1990). Models of Psychological Time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1–35). Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics, 59*(7), 1118–1140.

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

Cappelletti, M., Freeman, E. D., & Cipolotti, L. (2011). Numbers and time doubly dissociate. *Neuropsychologia, 49*(11), 3078–3092.

Champagne, J., & Fortin, C. (2008). Attention sharing during timing: Modulation by processing demands of an expected stimulus. *Perception & Psychophysics, 70*(4), 630–639. https://doi.org/10.3758/pp.70.4.630

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Eisler, H., Eisler, A. D., & Hellström, Å. (2008). Psychophysical issues in the study of time perception *Psychology of time*.(pp 75–109). Bingley: Emerald.

Ellis, N. (1992). Linguistic relativity revisited: The bilingual word-length effect in working memory during counting, remembering numbers, and mental calculation. *Advances in Psychology, 83*, 137–155.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Getty, D. J. (1976). Counting processes in human timing. *Perception & Psychophysics, 20*(3), 191–197. https://doi.org/10.3758/bf03198600

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*(3), 279–325.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences, 423*(May), 52–77.

Gilaie-Dotan, S., Rees, G., Butterworth, B., & Cappelletti, M. (2014). Impaired numerical ability affects supra-second time estimation. *Timing & Time Perception, 2*(2), 169–187.

Gilliland, A. R., & Martin, R. (1940). Some factors in estimating short time intervals. *Journal of Experimental Psychology, 27*(3), 243–255.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics, 72*(3), 561–582. https://doi.org/10.3758/app.72.3.561

Grondin, S., & Killeen, P. R. (2009a). Effects of singing and counting during successive interval productions. *Neuroquantology, 7*(1), 77-84.

Grondin, S., & Killeen, P. R. (2009b). Tracking time with song and count: Different Weber functions for musicians and nonmusicians. *Attention, Perception, & Psychophysics, 71*(7), 1649–1654.

Grondin, S., Laflamme, V., & Gontier, E. (2014). Effect on perceived duration and sensitivity to time when observing disgusted faces and disgusting mutilation pictures. *Attention, Perception & Psychophysics, 76*(6), 1522–1534. https://doi.org/10.3758/s13414-014-0682-7

Grondin, S., Laflamme, V., & Mioni, G. (2015). Do not count too slowly: Evidence for a temporal limitation in short-term memory. *Psychonomic Bulletin & Review, 22*(3), 863–868.

Grondin, S., Meilleur-Wells, G., & Lachance, R. (1999). When to start explicit counting in a time-intervals discrimination task: A critical point in the timing process of humans. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 993–1004.

Grondin, S., Ouellet, B., & Roussel, M. E. (2004). Benefits and limits of explicit counting for discriminating temporal intervals. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 58*(1), 1–12. https://doi.org/10.1037/h0087436

Hicks, R. E., & Allen, D. A. (1979a). Counting eliminates the repetition effect in judgments of temporal duration. *Acta Psychologica, 43*(5), 361–366. https://doi.org/10.1016/0001-6918(79)90030-1

Hicks, R. E., & Allen, D. A. (1979b). Repetition effect in judgments of temporal duration across minutes, days, and months. *American Journal of Psychology, 92*(2), 323–333. https://doi.org/10.2307/1421927

Hinton, S. C., & Rao, S. M. (2004). "One-thousand one... one-thousand two... ": Chronometric counting violates the scalar property in interval timing. *Psychonomic Bulletin & Review, 11*(1), 24–30. https://doi.org/10.3758/bf03206456

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Jones, M. R., & McAuley, J. D. (2005). Time judgments in global temporal contexts. *Attention, Perception, & Psychophysics, 67*(3), 398–417.

Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the weber function. *Psychological Review, 94*(4), 455–468. https://doi.org/10.1037//0033-295x.94.4.455

Mioni, G., Stablum, F., McClintock, S. M., & Grondin, S. (2014). Different methods for reproducing time, different results. *Attention, Perception, & Psychophysics, 76*(3), 675–681.

Rammsayer, T., & Ulrich, R. (2011). Elaborative rehearsal of nontemporal information interferes with temporal processing of durations in the range of seconds but not milliseconds. *Acta Psychologica, 137*(1), 127–133. https://doi.org/10.1016/j.actpsy.2011.03.010

Rattat, A. C., & Droit-Volet, S. (2012). What is the best and easiest method of preventing counting in different temporal tasks? *Behavior Research Methods, 44*(1), 67–80. https://doi.org/10.3758/s13428-011-0135-3

Ryan, L. J. (2011). Temporal context affects duration reproduction. *Journal of Cognitive Psychology, 23*(1), 157–170. https://doi.org/10.1080/20445911.2011.477812

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs, 77*(13), 1–31.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading: AddisonWesley.

Vierordt, K. v. (1868). *Der Zeitsinn nach Versuchen*. Tübingen: Laupp.

Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology. Animal Behavior Processes, 23*(4), 502–511. https://doi.org/10.1037//0097-7403.23.4.502

Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology (Hove), 61*(4), 569–587. https://doi.org/10.1080/17470210701282576

Woodrow, H. (1934). The temporal indifference interval determined by the method of mean error. *Journal of Experimental Psychology, 17*(2), 167–188.

Zakay, D., & Block, R. A. (1996). The role of attention in time estimation processes. In M. A. Pastor & J. Artieda (Eds.), Time, internal clocks, and movement (pp. 129–141). Amsterdam; New York: Elsevier.