

## University of Groningen

### Context Matters: Memories of Prior Times

Maaß, Sarah

DOI:  
[10.33612/diss.135934544](https://doi.org/10.33612/diss.135934544)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Maaß, S. (2020). *Context Matters: Memories of Prior Times*. University of Groningen.  
<https://doi.org/10.33612/diss.135934544>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## Chapter 2

### The Context Matters

This chapter has been published as: Maaß\*, S. C., Schlichting\*, N., & Van Rijn, H. (2019). Eliciting Contextual Temporal Calibration: The Effect of Bottom-up and Top-down Information in Reproduction Tasks. *Acta Psychologica 199*, 102898.

\*shared first author

## Abstract

Bayesian integration assumes that a current observation is integrated with previous observations. An example in the temporal domain is the central tendency effect: when a range of durations is presented, a regression towards the mean is observed. Furthermore, a context effect emerges if a partially overlapping lower and a higher range of durations is presented in a blocked design, with the overlapping durations pulled towards the mean duration of the block. We determine under which conditions this context effect is observed, and whether explicit cues strengthen the effect. Each block contained either two or three durations, with one duration present in both blocks. We provided either no information at the start of each block about the nature of that block, provided written (“short” / “long” or “A” / “B”) categorizations, or operationalized pitch (low vs high) to reflect the temporal context. We demonstrate that (1) the context effect emerges as long as sufficiently distinct durations are presented; (2) the effect is not modulated by explicit instructions or other cues; (3) just a single additional duration is sufficient to produce a context effect. Taken together, these results provide information on the most efficient operationalization to evoke the context effect, allowing for highly economical experimental designs, and highlights the automaticity by which priors are constructed.

# Introduction

When estimating the duration of a specific interval, we do not only perceive and process the duration of that specific interval, but also integrate other factors like prior knowledge about the statistical features of earlier perceived durations in the current environment. In fact, this has been proposed to serve as a way to optimize behavior, because the prior information can dampen the consequences of noise during the perception of a single event (Faisal, Selen, & Wolpert, 2008), a phenomenon especially relevant in noisy, real-world settings (Van Rijn, 2018). Evidence for the integration of sensory evidence and prior knowledge can be observed in many perceptual and cognitive tasks (e.g., for any magnitude estimation task, Martin, Wiener, & Van Wassenhove, 2017), and modeled with Bayesian observer models (e.g., Petzschner, Glasauer, & Stephan, 2015). Bayesian models of perception also have a significant impact on the field of time perception (for an overview, see Shi, Church, & Meck, 2013; Van Rijn, 2016), for example to explain the central tendency observed in multi-duration tasks, or temporal context effect.

The temporal context effect occurs when a particular interval is either over- or underestimated as a function of the distribution of the other test-intervals that are presented. More precisely, central tendency effects will cause an interval to be underestimated when presented alongside shorter intervals (i.e., within a temporal context of short intervals), while it will be overestimated when presented together with longer intervals (Jazayeri & Shadlen, 2010). In recent years, Bayesian observer models have been shown to accurately reproduce human behavior in timing tasks (Acerbi, Wolpert, & Vijayakumar, 2012; Cicchini, Arrighi, Cecchetti, Giusti, & Burr, 2012; Gu, Jurkowski, Lake, Malapani, & Meck, 2015; Jazayeri & Shadlen, 2010; Roach, McGraw, Whitaker, & Heron, 2017; Shi et al., 2013). In a Bayesian framework, the perceived duration of the current trial (the likelihood) is integrated with previously encountered intervals (the prior) to obtain a subjective percept (the posterior) which will be used for reproduction. The central tendency effect is explained by assuming that the mapping from likelihood to posterior results in a systematic over- or underestimation of intervals that are shorter or longer than the center of the prior distribution. When the same interval is present in two different contexts consisting of shorter or longer intervals, it will be underestimated in the short, and overestimated in the long context, yielding the context effect (see Jazayeri & Shadlen, 2010, Figure 2, for a visual depiction, or for an online simulation, <https://vanrijn.shinyapps.io/MaassVanMaanenVanRijn2019>).

Temporal reproduction can only be affected by different contexts if the presented stimuli or experimental setup gave rise to the creation of different

contexts. A straightforward approach is to temporally separate the different contexts, for example in different sessions (e.g., Jazayeri & Shadlen, 2010) or in different blocks within one experimental session (e.g., Roach et al., 2017). However, contexts can be presented intermixed, as long as trials from different contexts are otherwise dissociated. In a series of experiments Roach et al. (2017) presented intervals of different contexts intermixed (i.e., not blocked or otherwise temporally separated), while the different contexts were distinguishable by their associated physical properties (e.g., shorter intervals were presented on the left side of the screen, longer intervals on the right) or response mode (e.g., motor reproduction for shorter intervals, vocal reproduction for longer intervals). In other words, they explored how distinct stimuli of different contexts need to be in order to acquire distinct priors for each context. They found that if contexts are associated with different motor responses, participants form distinct priors within one session, but not if contexts are associated with different stimulus locations. Interestingly, after more extensive training (i.e., participating in multiple sessions), stimulus location was also shown to be an effective cue to dissociate contexts. Moreover, other work has demonstrated that even if different priors are built, they still influence each other (Taatgen & Van Rijn, 2011). These examples illustrate that context effects are very much dependent on seemingly minute details in the experimental design.

When physical properties of the stimuli are not indicative of the context, other aspects need to be sufficiently distinct between contexts to obtain a temporal context effect. In how far statistical properties of the stimulus material (i.e., the distributions of durations) have to differ between contexts has recently received some attention in literature (e.g., Roach et al., 2017; Acerbi, Wolpert, & Vijayakumar, 2012, Rhodes, Seth & Roseboom, 2018). For example, Rhodes et al. (2018) demonstrated that the behavior of human participants was best described by a Bayesian observer model that assumed a different prior based on the specific parameters of the presented signal (i.e., visual flashes or auditory tones, high or low pitch tones, or white noise versus pure tone audio). Another type of contextual information that can shape perception is that of symbolic or semantic cues. For example, Petzschner, Maier and Glasauer (2012) have shown that, in a linear displacement task, when trials from short and long contexts are presented intermixed and not in separate blocks, verbal cues (i.e., whether the stimulus on a particular trial will be drawn from the short or long context) causes behavioral effects similar to the typical context effect. Another example of how prior knowledge can affect behavioral performance from the time perception literature is the work of Dyjas, Bausenhardt, and Ulrich (2014) on stimulus order effects in interval comparison tasks. Typically, discrimination performance is lower if the comparison interval precedes the standard interval

(Bausenhart, Dyjas, & Ulrich, 2015; Lapid, Ulrich, & Rammsayer, 2008). However, when participants are informed about the stimulus order before a trial starts, this effect is greatly reduced suggesting that participants have, to a certain degree, top-down control on the processing of intervals (Dyjas, Bausenhart, & Ulrich, 2014). In the current work, we explore how bottom-up (i.e., statistical properties of the stimulus material) and top-down information (i.e., abstract knowledge about the experimental conditions) influence the buildup of temporal priors. This way, we aim to give recommendations under what conditions reliable context effects can be obtained efficiently.

Participants completed a single session interval reproduction task in which the intervals were represented by the duration of pure, continuous tones. Intervals in even blocks were sampled from a short temporal context, and in odd blocks from a longer context (or vice versa). In the baseline version of the experiment each context (short and long) was defined by three intervals: a standard medium (M) interval combined with two shorter intervals ( $S_2$  and  $S_1$ , short context), or the standard medium (M) interval with two longer intervals ( $L_1$  and  $L_2$ , long context). The M interval was the same in each context and condition (see also Figure 2). In Experiment 1A, we manipulated the distribution of the intervals within and between contexts to test whether and when participants do not form two priors anymore, but start to generalize across contexts. In a between-subjects design, we manipulated the width of theoretical prior-distributions within each temporal context (i.e., the spread of durations within the short and long context), as well as in the distance between the mean durations of the temporal contexts (i.e., the theoretical priors becoming more similar) as depicted in Figure 2.

In Experiment 1B we tested the effect of more or less abstract top-down categorization instructions on the temporal context effect, either by means of explicit information or by means of a manipulation of the physical properties of the presented interval (cf., Rhodes, Seth & Roseboom, 2018). Intervals of the short and long context were, as in Experiment 1A, presented in blocks. In a between-participant manipulation, participants received further information about the experiment in that they were either informed before the start of each block whether they would (1) perceive shorter or longer intervals; (2) whether they would perceive intervals from ‘set A’ or ‘set B’ without explicit information about the relative durations (but which were associated with shorter or longer intervals); (3) whether they will hear higher or lower pitched tones (which were, again, associated with shorter or longer intervals, but now this information was accessible on each trial, see also Rhodes et al. 2018); or (4) they received no further information at all (baseline condition). If participants integrate such additional information in their representations of the different

contexts, these representations could be more distinct from each other, or the experience of the previous block (i.e., the other context) could be disregarded more easily, leading to a quicker recalibration of the prior and, again, more distinct representations of the two temporal contexts. In either case, we would expect to find more pronounced context effects if participants receive additional information.

## Experiment 1A

### Methods

**Participants.** Seventy-six students enrolled in the Psychology program of the University of Groningen (Mage = 21.0 years, SDage = 2.67 years, range: 18-31 years; 52 females) completed the experiment in exchange for partial course credits. All subjects gave written informed consent to participate in the experimental protocol, approved by the Psychology Ethical Committee of the University of Groningen.

**Apparatus.** MacBooks (13", 2009) controlled all experimental events. Auditory stimuli were presented through headphones (Sennheiser, HD280 Pro), with volume individually adjusted to comfortable levels. The experiment was programmed using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) in Matlab R2014b.

**Participants.** Participants completed an interval reproduction task, consisting of eight blocks of 60 trials each. The blocks consisted of either shorter or longer intervals and were presented in turns, with the order counterbalanced by participant number. Each trial consisted of the presentation of an interval, and the reproduction of that interval (see Figure 1). Each trial commenced with an ITI of a random duration between 2 and 3 seconds sampled from a uniform distribution during which a fixation cross (“+”) was presented in the center of the screen. Then a “!” appeared on the screen for 700ms to prepare the subjects for the presentation of the interval. Following this, the interval was presented by means of a 440 Hz pure tone that lasted for the duration associated with the current trial. After completion of the tone, an interstimulus interval (ISI) of 1.5 seconds was presented while a “?” was displayed on screen. Then another 440 Hz pure tone was started, during which the “?” remained on the screen. Participants were instructed to press the spacebar when the earlier presented interval duration had passed.

To test the effect of temporal distribution, Experiment 1A consisted of four temporal distribution conditions differing in the intervals being presented in the short and long blocks (conditions are labelled as S2-S1-M-L1-L2, S2-M-L2, S1.5-M-L1.5 and S1-M-L1, see Figure 2). Intervals in temporal distribution

S2-S1-M-L1-L2 were spaced logarithmically, ranging from 625 (S2) to 1296 ms (L2). The letters are descriptive of the temporal distributions, for example, S2 corresponds to the shortest duration presented within distributions S2-S1-M-L1-L2 and S2-M-L2, and, L2 to the longest duration presented, S1.5 is the mean duration of S1 and S2, and, accordingly L1.5 is the mean duration of L1 and L2 (see Figure 2 for a graphical illustration). Intervals used in the other distributions were derived from this initial condition (S2-S1-M-L1-L2). In each block, each interval was presented twenty times while ensuring that each sequence of two durations appeared equally often. Stimulus order was determined using a de Bruijn sequence to ensure that every possible stimulus combination (i.e., duration of the previous trial and duration of the current trial) appeared equally often in the experiment (see also Wiener & Thompson, 2015). Because the S2-S1-M-L1-L2 distribution condition incorporated three intervals in each block type (short or long), each of the eight blocks consisted of 60 trials, while the other distribution conditions the eight blocks consisted of 40 trials.

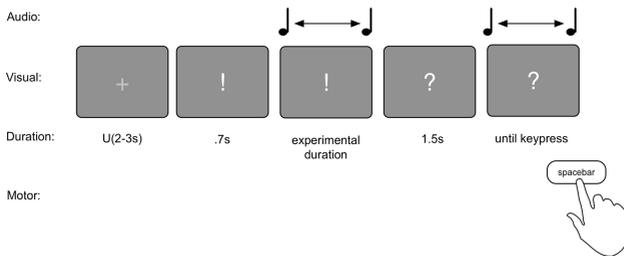
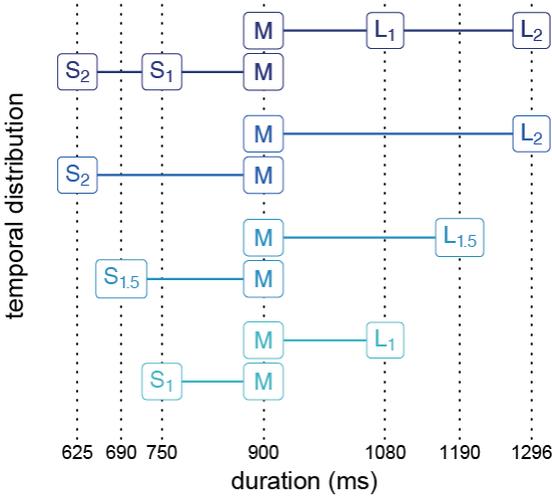


Figure 1. Graphical depiction of a single trial.

Data Analysis. A complete overview of all analyses and results can be found at <https://osf.io/m46v5>. Trials with a reproduction below 375 ms (40% shorter than the shortest interval) or longer than 1814 ms (40% longer than the longest interval) were discarded from the analysis. Three participants (out of 76) were excluded because more than 30% of the trials were to be discarded, for the remaining participants an average of 0.72% of trials were removed from subsequent analyses. All data were modeled using Linear Mixed Models (LMMs) using the lme4 (Bates, Mächler, Bolker, & Walker, 2014) and lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) in R (R Core Team, 2016). In all models, participant was entered as random intercept. Random slope terms were included if they improved the model based on likelihood ratio tests.



**Figure 2. Between subject conditions.** Each participant was assigned to one of four *temporal distribution* conditions (color coded, from top to bottom: S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub>, S<sub>2</sub>-M-L<sub>2</sub>, S<sub>1.5</sub>-M-L<sub>1.5</sub> and S<sub>1</sub>-M-L<sub>1</sub>). Each condition consisted of a short (left) and long (right) context blocks, sharing the “M” duration. Conditions differed in the spread of durations within contexts (e.g., distribution S<sub>1</sub>-M-L<sub>1</sub> is more narrowly spread than distribution S<sub>2</sub>-M-L<sub>2</sub>), as well as in the distance between the mean durations of temporal contexts (e.g., the difference between means of short and long context in the S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub> distribution is 292 ms versus 165 ms in the S<sub>1</sub>-M-L<sub>1</sub> distribution).

First, we tested the effect of temporal distribution on two different measures of context effects. The first measure is the M interval effect, here defined as the difference in reproductions of the M interval dependent on context. These M interval effects can be a result of regression towards the mean effects, but also of other factors that potentially influence intercepts within context conditions, like temporal distribution or categorization. We created an LMM to predict reproductions of the M interval, entering *context* (coded as -0.5 for context short and 0.5 for context long) and *temporal distribution*, as well as their interaction as predictors. We explicitly tested a model including the interaction term *context* × *temporal distribution* but not *temporal distribution* as a main effect, because *temporal distribution* could have differential effects on M interval reproductions dependent on context, but not a general effect. We compared these models by quantifying the evidence, denoted as BF01, in favor of the null hypothesis (H0, here expressed by the less complex model) over the alternative hypothesis (H1, here expressed by the more complex model) using the Bayesian Information Criterion (BIC) calculated for both models (Wagenmakers, 2007).

As a second measure we compared slope values obtained from linear regressions that were performed for each context and participant separately. These values reflect the regression towards the mean within each context (i.e., slope values close to 1 can be interpreted as almost no regression towards the mean, and lower slope values as more regression towards the mean). *Slope* was entered in the LMM as the dependent variable, while again *context* and *temporal distribution* including their interaction were entered as predictors. We then proceeded as described above to test the effect of temporal distribution.

Additionally, we tested whether sequential context effects (i.e., the effect of the previous presented duration on the current reproduction) differed between temporal distribution conditions. The initial model included *reproduction* as dependent variable; *duration* (i.e., presented duration; entered as a continuous value, centered on the M interval), *temporal distribution*, and *context* as fixed factors; and the interaction terms *duration*  $\times$  *temporal distribution* and *context*  $\times$  *temporal distribution* (cf., Taatgen & Van Rijn, 2011, Hallez, Damsma, Rhodes, Van Rijn, & Droit-Volet, 2019). We sequentially added previous presented durations (*N-1*, *N-2*, etc.) and interaction terms *N-1* (*N-2*, etc.)  $\times$  *temporal distribution* and tested whether they improved the model fit as described above.

**Table 1. Summary of model analyses for Experiment 1A.** Rows indicate different models, ordered from complex models including all interactions (Roman numeral IV) to simple main effect models (Roman numeral I). Bayesian Information Criterion (BIC) as computed by the lme4 package. Bayes Factors (BF) express the evidence in favor of the model associated with the row, against the model associated with the column (indicated by Roman numbers). Green shades indicate comparisons in favor (BF>10) of the model associated with the row, red shades indicate comparisons in favor of the model associated with the column (BF<0.1). (“baseline model”: “duration  $\times$  distribution + context  $\times$  distribution”, “distr.”: distribution).

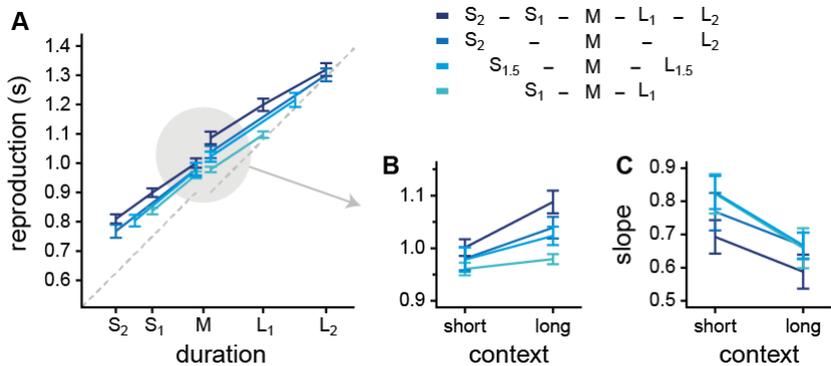
	BIC	BF <sub>column,row</sub>	IV	III	II	I
<b>reproduction M interval ~</b>						
IV	context $\times$ distribution	-11430		1.2x10 <sup>-4</sup>	5.3x10 <sup>+11</sup>	3.9x10 <sup>+7</sup>
<b>III</b>	<b>context + context : distribution</b>	<b>-11448</b>	<b>8.1x10<sup>+3</sup></b>		<b>4.3x10<sup>+15</sup></b>	<b>3.2x10<sup>+11</sup></b>
II	context + distribution	-11376	1.8x10 <sup>+12</sup>	2.3x10 <sup>-16</sup>		7.4x10 <sup>-5</sup>
I	context	-11395	2.5x10 <sup>-8</sup>	3.1x10 <sup>-12</sup>	1.3x10 <sup>+4</sup>	
<b>slope ~</b>						
IV	context $\times$ distribution	-2		0.007	9.1x10 <sup>-4</sup>	3.7x10 <sup>-6</sup>
III	context + context : distribution	-12	148.413		0.135	5.5x10 <sup>-4</sup>
II	context + distribution	-16	1.1x10 <sup>+3</sup>	7.389		0.004
<b>I</b>	<b>context</b>	<b>-27</b>	<b>2.6x10<sup>+5</sup></b>	<b>1.8x10<sup>+3</sup></b>	<b>244.692</b>	
<b>reproduction ~</b>						
IV	+ N-1 $\times$ distr. + N-2 $\times$ distr.	-24805		6.2x10 <sup>-11</sup>	1.6x10 <sup>-10</sup>	1.1x10 <sup>+11</sup>
<b>III</b>	<b>+ N-1 + N-2</b>	<b>-24852</b>	<b>1.6x10<sup>+10</sup></b>		2.718	<b>1.9x10<sup>+21</sup></b>
II	+ N-1	-24850	5.9x10 <sup>+9</sup>	0.368		7.0x10 <sup>+20</sup>
I	+ $\emptyset$ (only baseline model)	-24754	5.2x10 <sup>-22</sup>	5.2x10 <sup>-22</sup>	1.4x10 <sup>-21</sup>	

## Results

Results of Experiment 1A are summarized in Figure 3 and in Table 1. Model comparisons testing the effect of temporal distribution on reproductions of the M interval (Figure 3B) are shown in the top section of Table 1. The preferred model (Model III) includes a main effect of *context* and the interaction between *context* and *temporal distribution*. Note that *temporal distribution* was not included as a main effect in this model to test the assumption that *temporal distribution* in itself does not influence the difference in reproductions of the M interval. *Context* influences M interval reproductions ( $\beta = 0.09$ ,  $t(74.08) = 9.09$ ,  $p < .001$ , given the S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub> distribution), an effect that is modulated by temporal distribution (S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub> vs. S<sub>1.5</sub>-M-L<sub>1.5</sub>  $\beta = -0.04$ ,  $z = -3.04$ ,  $p = .013$ , S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub> vs. S<sub>2</sub>-M-L<sub>2</sub>  $\beta = -0.07$ ,  $z = -4.99$ ,  $p < .001$ , and S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub> vs. S<sub>1</sub>-M-L<sub>1</sub>  $\beta = -0.04$ ,  $z = -2.88$ ,  $p = .020$ ). These effects are reflected in Figure 4B, with steeper lines reflecting stronger context effects.

Comparison of different models predicting slope values prefers a model with just a main effect of *context* (Table 1, middle section, Model I), without additional main effects or interactions of temporal distribution. The *context* effect reflects lower slope values in the long compared to the short context ( $\beta = -0.13$ ,  $t(74) = -5.03$ ,  $p < .001$ ; see also Figure 3C).

The analysis of sequential context effects (see the model comparisons in the lower section of Table 1) revealed that both N-1 and N-2 affect the current reproduction ( $\beta = 0.07$ ,  $t(26467) = 10.31$ ,  $p < .001$ ; and  $\beta = 0.02$ ,  $t(26467) = 3.71$ ,  $p < .001$ , respectively) as Model III was preferred. Note that there is limited evidence in favor of Model III over Model II (BF=2.71), so alternatively one could have selected Model II. However, as both models are qualitatively similar and evidence does prefer Model III over Model II, we have opted for Model III. (Note that a comparison with a model also including N-3 preferred the simpler model, BF=0.006). In summary, reproductions were influenced by previously presented durations (with strong evidence for N-1, and reduced evidence for N-2).



**Figure 3. Effect of temporal distributions on context effects.** A, Average interval reproductions dependent on context and temporal distribution. Note that the overlapping interval M has the same duration (0.9 s) in both contexts, but is spatially separated dependent on context for display purposes. B, Reproductions of the M interval dependent on context and temporal distribution. C, Average slope values dependent on context and temporal distribution. Error bars depict the standard error.

## Experiment 1B

### Methods

**Participants.** Eighty-one students enrolled in the Psychology program of the University of Groningen (Mage = 20.6, SDage = 2.8 years, range: 17-33 years; 58 females) completed the experiment in exchange for course credits. All subjects gave written informed consent to participate in the experimental protocol, approved by the Psychology Ethical Committee of the University of Groningen.

**Apparatus.** Identical to Experiment 1A (see above).

**Procedure.** Participants completed the same reproduction task as in Experiment 1A (see above). The intervals presented were identical to the durations of temporal distribution S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub> of Experiment 1A (see Figure 2).

To manipulate the strength of categorization, participants were assigned to one of four versions of Experiment 1B which differed in the information that was given about the blocks in form of written instructions prior the beginning of the task. In the version *no label* no further instructions were given about the lengths of intervals in each block (note that this is the same group of participants as in Experiment 1A, temporal distribution S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub>). In the version *pitch* the following instructions were given prior to the start of the task “In

the next blocks, you will hear durations with different pitches. The pitch changes from block to block”. In the version *short/long* instructions read “In the next blocks, you will either hear SHORTer or LONGer durations. This will be indicated at the start of each block.” with a further instruction indicating whether the lengths of intervals in the next block are short or long. In the version *A/B* participants were informed that “In the next blocks, you will hear durations from SET A or SET B. This will be indicated at the start of each block.” with the corresponding set denoted before each block.

The rest of the procedure in Experiment 1B was identical to condition  $S_2$ - $S_1$ - $M$ - $L_1$ - $L_2$  of Experiment 1A with the exception that in version *pitch* the intervals were presented with a 420 Hz pure tone in the short/long blocks and a 460 Hz pure tone in the long/short blocks. Whether the short block was associated with the higher or lower pitch tone was counterbalanced between participants.

**Data Analysis.** Identical rejection criteria were used as in Experiment 1A (see 2.1.4.). Trials with a reproduction below 375 ms (40% shorter than the shortest interval) or longer than 1814 ms (40% longer than the longest interval) were discarded from the analysis. One participant was excluded because more than 30% of the trials were to be discarded, for the remaining participants an average of 2.04% of trials were removed from subsequent analyses. Data of Experiment 1B was analyzed using the same analysis pipeline as in Experiment 1A, with the only difference that instead of including temporal distribution as secondary predictor, here we included *categorization* as a predictor in the model analyses.

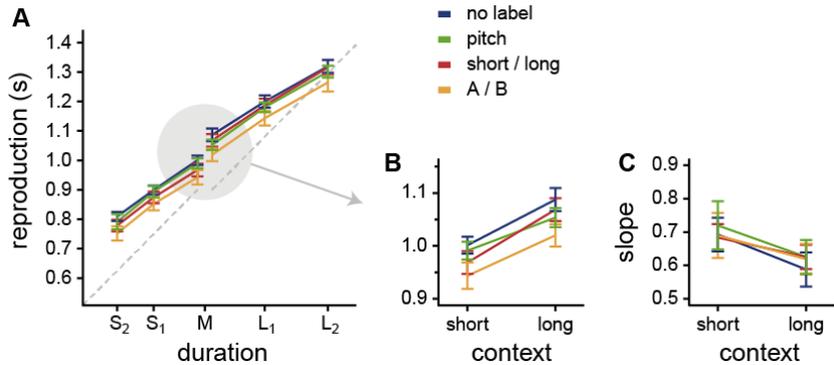
## Results

In Experiment 1B we tested the effect of categorization on the strength of context effects (Table 2, top section, and visually depicted in Figure 4A and 4B). Model analysis of reproductions of the M interval revealed that there was no evidence in favor of the inclusion of *categorization* as a main factor or an interaction including *categorization*. The preferred Model I contains a *context* factor independent of categorization condition: participants tended to overestimate the M interval in the long compared to the short context ( $\beta = 0.08$ ,  $t(79.04) = 15.33$ ,  $p < .001$ ).

We found the same pattern of results when testing slope-values (Table 2, middle section, Figure 4C). Again, based on model comparisons we find no evidence for an effect of categorization on slope, but did find a main effect of context in general, meaning that in the long context (lower slope values) par-

Participants relied more on prior representations than in the short context (higher slope values,  $\beta = -0.08$ ,  $t(80) = -3.75$ ,  $p < .001$ ).

Similar to Experiment 1A, the analysis of sequential context effects revealed that both  $N-1$  and  $N-2$  affect the current reproduction ( $\beta = 0.07$ ,  $t(37456) = 11.23$ ,  $p < .001$ ; and  $\beta = 0.03$ ,  $t(37456) = 5.43$ ,  $p < .001$ , respectively, based on the model including  $N-1$  and  $N-2$  but not their interaction with *categorization*), but not  $N-3$  ( $BF = 0.01$ ).



**Figure 4. Effect of categorization on context effects.** Note that the condition *no label* is identical to the condition S<sub>2</sub>-S<sub>1</sub>-M-L<sub>1</sub>-L<sub>2</sub>, in Experiment 1A (see Figure 3), in both cases color coded as dark blue (baseline condition). **A**, Average interval reproductions dependent on context and categorization condition. Note that the overlapping interval M has the same duration (0.9 s) in both contexts, but is spatially separated dependent on context for visual clarity. **B**, Reproductions of the M interval dependent on context and categorization condition. **C**, Average slope values dependent on context and categorization condition.

**Table 2. Summary of model analyses for Experiment 1B.** Rows indicate different models, ordered from complex models including all interactions (Roman numeral IV) to simple main effect models (Roman numeral I). Bayesian Information Criterion (BIC) as computed by the lme4 package. Bayes Factors (BF) express the evidence in favor of the model associated with the row, against the model associated with the column (indicated by Roman numbers). Green shades indicate comparisons in favor (BF>10) of the model associated with the row, red shades the model associated with the column (BF<0.1). (“baseline model”: “duration × categorization + context × categorization”, “cat.”: categorization).

	BIC	BF <sub>column,row</sub>	III	II	I
<b>reproduction M ~</b>		IV			
IV context × categorization	-9080		1.0x10 <sup>-5</sup>	0.0302	3.0x10 <sup>-7</sup>
III context + context : categorization	-9103	9.8x10 <sup>+4</sup>		2.9x10 <sup>+3</sup>	0.0302
II context + categorization	-9087	33.1155	3.3x10 <sup>-4</sup>		1.0x10 <sup>-5</sup>
<b>I context</b>	<b>-9110</b>	<b>3.2x10<sup>+6</sup></b>	<b>33.1155</b>	<b>9.8x10<sup>+4</sup></b>	
<b>slope ~</b>					
IV context × categorization	-4		5.5x10 <sup>-4</sup>	9.1x10 <sup>-4</sup>	5.0x10 <sup>-7</sup>
III context + context : categorization	-19	1.8x10 <sup>+3</sup>		1.6487	9.1x10 <sup>-4</sup>
II context + categorization	-18	1.1x10 <sup>+3</sup>	0.6065		5.5x10 <sup>-4</sup>
<b>I context</b>	<b>-33</b>	<b>1.9x10<sup>+6</sup></b>	<b>1.1x10<sup>+3</sup></b>	<b>1.8x10<sup>+3</sup></b>	
<b>reproduction ~ baseline model</b>					
IV + N-1 × cat. + N-2 × cat.	-26488		2.2x10 <sup>-11</sup>	1.8x10 <sup>-7</sup>	7.7x10 <sup>+18</sup>
<b>III + N-1 + N-2</b>	<b>-26537</b>	<b>4.3x10<sup>+10</sup></b>		<b>8.1x10<sup>+3</sup></b>	<b>3.4x10<sup>+29</sup></b>
II + N-1	-26519	5.3x10 <sup>+6</sup>	1.2x10 <sup>-4</sup>		4.2x10 <sup>+25</sup>
I + ∅ (only baseline mode)	-26401	1.2x10 <sup>-19</sup>	2.9x10 <sup>-30</sup>	2.3x10 <sup>-26</sup>	

## Discussion

When intervals are embedded in a temporal context, humans form an internal representation of the statistics of this context, or, in Bayesian terms, a prior distribution. If two or more temporal contexts are presented within the same experiment, participants can, under certain circumstances, form multiple priors – one for each temporal context – that differentially affect behavior. Presenting a standard interval (medium, M) either embedded into a short or a long context in a blocked design leads to the context effect: central tendency effects arise, causing the subjective estimations for the identical M interval to differ between contexts (i.e., relative underestimation in the short, and overestimation in the long context). Following this Bayesian rationale, the context effect can only be observed when the presented stimuli give rise to independent prior distributions, or, in other words, in order to obtain a temporal context effect, the contexts need to be perceived as sufficiently distinct from each other. Here, we explored the parameters that promote the formation of different priors by manipulating the underlying statistical properties in terms of the distribution of durations in each context (i.e., bottom-up information, Experiment 1A), or by

explicit instructions aiding the observers to categorize intervals into two distinct contexts (i.e., top-down information, Experiment 1B). Results show that if the means of two interval distributions are sufficiently different from each other, a separate central tendency effect emerges for each of the distributions, even in absence of explicit instructions. If the durations from both temporal distributions are too similar, however, participants will generalize over all presented durations, irrespective of the temporal context. Adding additional cues that potentially help participants to distinguish between temporal contexts did not improve the separation of both sets of intervals. These results are interesting as earlier work (e.g., Rhodes et al., 2018) has shown that unique prior distributions can be formed when intervals are defined by physical properties of the signal. This work indicates that presenting “blocks-of-trials” is sufficient to generate unique prior distributions, and adding additional cues to aid categorization did not affect the prior distributions sufficiently to be reflected in central tendency related statistics in this study.

Whenever central tendency effects were observed, we replicated the finding that the central tendency effect is increased in longer contexts (e.g., Jazayeri & Shadlen, 2010; Acerbi, Wolpert & Vijayakumar, 2012). This directly follows from the interaction between scalar property-driven modulations of precision as a function of duration and the influence of precision on the weighting of prior knowledge: Scalar timing entails that longer durations are perceived more noisily, causing wider likelihoods, and thus a stronger relative influence of the prior than with shorter, more accurately perceived durations (e.g., Bausenhart, Dyjas, & Ulrich, 2014; Wearden, 1991). In both experiments we found sequential context effects (N-1 and N-2) that were not influenced by the experimental manipulations, demonstrating the robustness of these effects (cf., Van Rijn, 2016).

To conclude, we have demonstrated that (1) participants implicitly distinguish between different contexts as long as the contexts contain sufficiently different durations, and that (2) neither explicit instructions, nor manipulations of physical aspects of the stimuli influence the magnitude of context effects when contrasted against a typical baseline condition. Below, we will first discuss how different distributions of durations influence the observed central tendency, and then discuss the top-down manipulations to increase context distinctiveness, followed by a discussion of factors influencing the observed results.

## Bottom-up Influences

Manipulating the statistical properties of the stimulus sets, that is, the distribution of durations presented in Experiment 1A (Figure 2), resulted in the following pattern of results: The largest context effect, as measured by the larg-

est absolute difference between the standard duration (medium, M) in the short and long context, was observed in conditions with the broadest range of intervals (spanning from S2 to L2). The temporal distribution S1.5-M-L1.5 was narrower and even though we still found a context effect, it was decreased as compared to wider temporal distributions (i.e., those that included S2 and L2). Thus, as a context effect was observed in these conditions (S2-S1-M-L1-L2, S2-M-L2, and S1.5-M-L1.5), distinct priors were formed for each temporal context. Even though comparisons need to be made carefully because of differences in experimental design, a similar pattern of adaptation to the statistical properties of the presented durations was observed in the work of Acerbi, Wolpert and Vijayakumar (2012).

In condition S1-M-L1, in which the narrowest temporal distribution was presented, participants generalized over both temporal contexts. That is, there is no evidence for the classification of the intervals into two separate contexts, as evidenced by the absence of a context effect (Figure 3B). Hence, two intervals per context are only sufficient to create two separate prior distributions if the longest/shortest interval has a duration that is sufficiently different from the other context. Otherwise, the three durations are perceived as originating from one temporal distribution and only a single prior will be formed, resulting in an overall central tendency effect across all durations (note that the pull of the longer durations will still be larger than the pull of the shorter duration due to the scalar timing property, Figure 3C).

As interval timing is an inherently noisy process, it is not surprising that a certain distance is needed to illicit the creation of two prior distributions. At the start of the experiment, participants build a prior that is constructed on the basis of the two or three durations presented in that block. If the objective durations presented in the subsequent block are relatively similar to the durations presented in the first block, the subjective durations - which are influenced by both noise and the prior that was constructed earlier - might be insufficiently different from the existing prior to be perceived as potentially resulting from a different context. In such cases, the participant will implicitly assume that all intervals are sampled for a single context, and thus no information is present that would warrant the formation of a new prior. This rationale is in line with memory theories that assume that an assessment of the content of the memory trace is used to determine in which context a trace is stored (e.g., Polyn, Norman, Kahana, 2009). However, this does raise the question of what the actual distance is that is necessary to form two distinct priors dependent on temporal context. We can express the relative distances as the longest interval divided by the standard interval (medium, M), and conclude that with a relative distance of 1.3 (L1.5/M), two priors are formed, but with a relative distance of 1.2

(L1/M) stimuli are perceived as coming from one theoretical prior distribution. It is noteworthy, however, that the emerging context effect is influenced not only by the distribution of the theoretical prior but also by the internal variability associated with the perception of the current duration (i.e., the likelihood). This internal variability, even in a very homogeneous population as tested here, has been shown to vary between participants and to influence context effects (Maaß & Van Rijn, 2018). The reasoning outlined above creates an interesting paradox. If sampling would have resulted in a subset of participants with very low internal variability (cf., Cicchini et al., 2012), we would, on the one hand, have expected a smaller or even absent context effect in the conditions in which we now found context effects. Yet, due to their higher precision, these participants are more likely to notice that a new block is associated with a different distribution even when the differences between durations are relatively small, which could have resulted in a reliable context effect in the condition in which we now did not observe any effect. Similar reasoning also holds when considering the parameterization of the task used to measure the context effects. For example, in the work reported in this study, we remained close to our earlier studies in which we asked participants to end a machine-initiated interval by a single keypress (e.g., Schlichting et al., 2018; Van Rijn & Taatgen, 2008). However, it has been demonstrated that the method used for reproducing durations influenced both accuracy and precision, with an offset-keypress associated with lower precision than having the participant keep a key depressed for the duration of the reproduction (Mioni, Stablum, McClintock, & Gronin, 2014, see also Damsma, Schlichting, Van Rijn, & Roseboom, 2019). Thus, if we would have utilized a reproduction paradigm in which participants were asked to keep a key depressed for the perceived duration, we might have observed a reliable context effect in condition S1-M-L1. Similar reasoning could be applied to whether filled (more precise) or open (less precise) intervals (e.g., Grondin, 2010; Rammseyer & Altenmueller, 2006) or auditory (more precise) or visual (less precise) stimuli are used (e.g., Aagten-Murphy, Cappagli, & Burr, 2014; Burr, Banks, & Morrone, 2009). How such paradigmatic differences influence the central tendency effects could be tested in future research.

Summarizing, higher or lower precision – either due to different participant characteristics or different task design – will influence the observed results. Our results show that just a single additional duration presented in a block with a standard duration suffices to elicit a reliable context effect, and given the current parameterization of the task and the population tested, the relative distance of this additional duration should be in the order of 1.3 times the standard duration.

## Top-down Influences

As a second type of manipulation, we provided the participants with additional cues that could potentially aid in distinguishing temporal contexts. Interestingly, providing participants with information that would allow for easier distinction of the temporal context did not result in a stronger central tendency or context effect. Information about the nature of the block was either presented before each block commenced or by varying the physical properties of the stimulus itself. The written instructions before each block either indicated this block to be “short” or “long”, or used a more neutral labeling (i.e., block “A” or “B”). In another version, we did not present participants with written instructions before each block, but presented the duration in different pitch as a function of context (as in Rhodes et al., 2018). Yet, even changing the properties of the stimulus itself did not result in a stronger context effect, even though the disambiguating information was available on each trial as compared to at the start of the block. In line with findings by Roach et al. (2017), these results suggest that duration is stored independently from other physical stimulus attributes.

It is possible that we failed to find an effect because the implicit emergent context effect observed in the baseline condition was already driven by participants forming two separate priors. Adding explicit cues might have had an effect if durations from different blocks were perceived as coming from one theoretical prior distribution, as we observed in Experiment 1A, condition S1-M-L1. This is supported by the work of Petzschner, Maier and Glasauer (2012), who demonstrated that with explicit cues two priors could be formed in an experiment that elicited a single prior without cues. However, as the baseline version was manipulated in this study, this question remains to be addressed in future work. Nevertheless, it is surprising that Version 2, in which the blocks were labeled explicitly as “Short” and “Long”, did not result in stronger under- or overestimation, as irrespective of the current trial one might be primed towards producing shorter or longer intervals.

Taken together, the two sets of experiments show that as little as 40 trials per block are sufficient to build a prior, and that a context effect is observed after a single pair of a short and long block (see Figure S1, available at <https://osf.io/m46v5>). Moreover, over the course of the experiment, participants are able to rapidly adapt to varying temporal contexts. This also happens if no information is given about stimuli belonging to a different set, but pivotal to that is that there are at least some notably distinguishably different durations presented (as can be seen by the lack of context effects in Experiment 1A, condition S1-M-L1.).

## Conclusion

The aim of this study was to gain deeper insight into the acquisition of reliable context effects in timing tasks, and the factors influencing the formation of different priors for different temporal contexts. To this end, we investigated how on the one hand bottom-up (i.e., statistical properties of the stimulus material) and on the other hand top-down information (i.e., more abstract or more explicit cues about the experimental conditions) calibrate the temporal context. Interestingly, providing information on the distinctness of the different contexts did not have any effect on an observed context effect, indicating that top-down information does not strengthen the build-up of distinct priors. However, the width of theoretical prior-distribution does determine how instances of different magnitudes are clustered. Thus, context effects are very robust to top-down information, with contexts not needing to be explicitly differentiated in block instruction as long as there is a noticeable difference between the distributions of durations used.

To summarize, when reliable context effects are needed, the presentation of just a single duration in addition to a standard duration suffices, without requiring explicit instructions. This is valuable information when designing experiments in which the under- or overestimation of a standard duration is focus. By just presenting a single shorter or a single longer duration in addition to a standard duration, one can optimize the number of critical trials, and therefore either keep experiments relatively short (highly relevant in, for example, children or clinical populations studies, e.g., Hallez et al., 2019, Karaminis et al., 2016) or optimize the number of critical trials in, for example, decoding studies.

