A Process Model of Magnitude Estimation

J. Gregory Trafton (greg.trafton@nrl.navy.mil)
Naval Research Laboratory
Washington, DC 20375 USA

Abstract

We present a cognitively plausible model of non-verbal counting and magnitude estimation. Unlike existing models, the current model does not use a perfect representation of magnitude, time, or memory. Instead, it calculates a magnitude based on an imperfect rate of counting and determines when to stop counting based on an internal timer. Empirical data at both the individual and average level is matched to show a range of performance.

Keywords: magnitude-estimation; nonverbal counting; cognitive modeling

Introduction

Numerosity, magnitude, estimation, and counting are fundamental aspects to human life. Some researchers have suggested that numerosity is one of people's core concepts (Carey, 2009). Other researchers have shown that animals can count, even without an explicit (verbal) counting mechanism (Platt & Johnson, 1971).

In fact, magnitude has been explored intensively using a variety of methods including counting (Whalen, Gallistel, & Gelman, 1999), size (Moyer & Landauer, 1967), math (C. Gallistel & Leon, 1991; C. R. Gallistel & Gelman, 2000), and perception of number (Wynn, 1992). Counting, while being one of the purest measures of magnitude, has probably been studied the least, at least in humans. Additionally, there are relatively few cognitive models of counting, though there are other models of size, perception and math.

Here, we are concerned with *non-verbal counting*, where a person performs an action (e.g., lever-pressing) a specified number of times without explicit enumeration (Whalen et al., 1999). Mathematicians and philosophers have argued that non-verbal counting is the basis of higher-order math (Bell, 1937). Non-verbal counting is also used in many everyday situations, from determining how much trash is in a garage to the number of people in a queue to the number of steps on an escalator.

Previous researchers have suggested that non-verbal counting occurs through an internal, noisy accumulator (Meck & Williams, 1997; Gibbon, Church, & Meck, 1984; Meck & Church, 1983). In these accounts, magnitudes have scalar variability, varying in proportion to the mean of the magnitude (C. R. Gallistel & Gelman, 2000). Because magnitudes have scalar variability, the discriminability of the values obeys Weber's law because the degree of overlap between representations remains constant as the ratio of the means is held constant. Current accounts make these assumptions:

- An accumulator is incremented based on count (Cordes, Gelman, Gallistel, & Whalen, 2001; Whalen et al., 1999) or time (Dormal, Seron, & Pesenti, 2006; Meck & Williams, 1997).

- The accumulator value has a perfect representation, but when checked internally is noisy; the bigger the value of the accumulator, the bigger the noise (C. R. Gallistel & Gelman, 2000; Meck & Church, 1983; Meck, Church, & Gibbon, 1985).

There are several major concerns with these assumptions, however.

Over-reliance on perfect accumulators or perfect memory: First, most accounts that assume that the accumulator is based on an actual count assume that the counter is perfect (Cordes et al., 2001; Whalen et al., 1999), which is cognitively implausible. In these approaches, the counter is represented perfectly, but is retrieved with noise. For accounts that assume that the accumulator is based on an internal timer, the assumption is that the timer is perfect (Meck et al., 1985). We know from many studies of time sense that people do not have perfect representations of time (Zakay & Block, 1997; Matell & Meck, 2000) and that people are able to estimate time more accurately at shorter intervals than longer intervals. At least some of these approaches also assume perfect memory (Gibbon et al., 1984). These assumptions allowed early progress to be made on the initial models and theorizing, which clearly advanced the field. Unfortunately, these assumptions have continued on through many of the current models of counting and may lead to an incorrect understanding of how people perform non-verbal counting.

Sampling problem: If a human counter samples magnitude from a Gaussian distribution and periodically checks that magnitude against a target goal, a trace of the magnitude across a counting scenario will show it to sometimes become negative or go backwards (a standard assumption of most accumulator models and inherent in consecutive random sampling). If a further constraint is added so that the magnitude must be positive and always increase, the magnitude will consistently *under* represent the actual count. This under counting will become greater the bigger the target is because there is more opportunity for skipping a number.

Our goal here is to remove these problems and present a process model of how people perform these implicit counting tasks. We assume that people do not have a perfect sense of memory, time, or magnitude when counting non-verbally. We describe our model in the context of a classic counting experiment by Whalen et al. (1999).

Method (Whalen et al., 1999)

A complete description of the experiment can be found in Whalen et al. (1999).

Participants

Seven volunteers participated in the experiment over 8 1-hour sessions (which included other related tasks as well).

Setup and Procedure

A trial began with a ”Ready?” message in the center of the screen. When the participant pushed a button, the ”Ready?” message was replaced with an odd number from 7 - 25 (inclusive). Participants were instructed to push a key the specified number of times, as fast as they could. Participants completed a trial by pushing a different key. Participants performed 40 trials for each odd number from 7 - 25. No feedback was given regarding their accuracy.

Participants were specifically instructed **not** to verbally count the number of presses made, but to arrive at their target-goal “by feel.”

Measures

The target-goal and the number of actual keypresses was recorded and averaged for each participant. The standard deviation and coefficient of variation was also measured for each participant.

Results and Discussion

Participants were reasonably accurate for most target-goals. The average number of presses increased linearly with the target value. For all participants, the standard deviation of the number of key presses varied in direct proportion to the target magnitude.

The most surprising finding, however, concerned the coefficient of variation (the ratio of standard deviation and mean). Specifically, the coefficient of variation was *constant* across target size. Figure 1 shows the averaged data across the seven participants (digitally extracted from the original article).

Participants were presumably not performing overt or covert verbal counting because the rate that they were able to push the key (∼120ms/item), is much faster than subvocal counting can occur (∼240ms; Klahr, 1973). In fact, when participants were instructed to explicitly subvocalize, their RT was significantly and consistently longer than when they performed the non-verbal counting task. The difference between subvocalizing and non-verbal counting was much bigger when the numbers had more syllables (e.g., ”nine” vs. ”seventeen”).

Architecture and Model Description

ACT-R is a hybrid symbolic/sub-symbolic production-based system (Anderson et al., 2004) ACT-R consists of a number of modules, buffers, and a central pattern matcher. Modules in ACT-R contain a relatively specific cognitive faculty usually associated with a specific region of the brain. For each module, there are one or more buffers that communicate directly

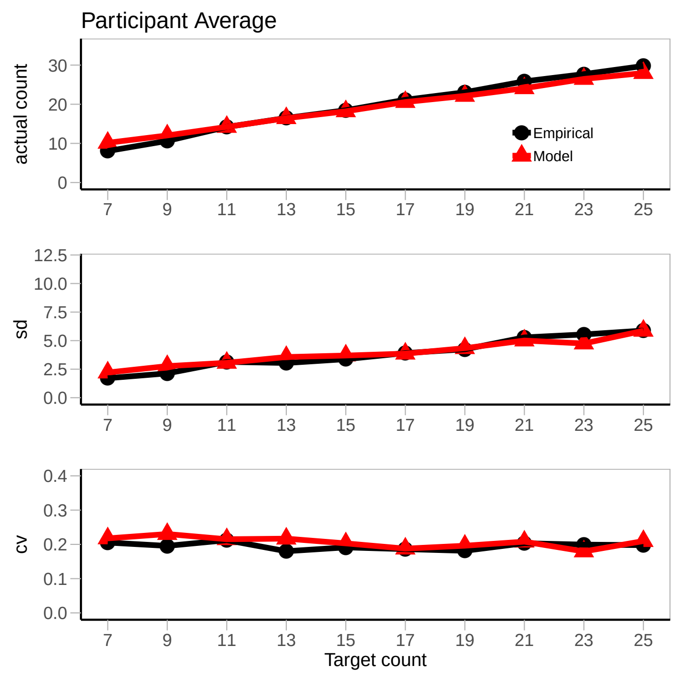


Figure 1: Average performance of the seven individuals in the Whalen et al. (1999) study. The x axis in all three graphs is the target count (the goal that the participants were given). The top panel shows the (remarkably accurate) accuracy on counting. The middle panel shows the increasing standard deviation the higher the target goal becomes. The bottom panel shows the flat coefficient of variation. The darker circles show the data digitally extracted from the original article while the lighter triangles show the model fit.

with that module as an interface to the rest of ACT-R. At any point in time, there may be at most one item in any individual buffer; thus, the module's job is to decide what and when to put a symbolic object into a buffer. The pattern matcher uses the contents of the buffer to match specific productions.

ACT-R uses if-then rules (productions) that will fire when their preconditions are met by matching the contents of the buffers. If there is more than one production that can fire, the one with the highest utility (production strength) will fire. Each production can change either internal state (e.g., buffer contents) or perform an action (e.g., click on a button).

ACT-R interfaces with the outside world through the visual module, the aural module, the motor module, and the vocal module. The architecture supports other faculties through intentional, imaginal, temporal and declarative modules.

Because most researchers believe that numerosity is a core concept (Carey, 2009) and many animals can actually count non-verbally, we have created a new ACT-R module, called the magnitude module.

The Magnitude Module

The magnitude module provides a mechanism for performing non-verbal counting until a specific target-goal is reached.

Instead of relying on a perfect counter or a perfect sense of time, the magnitude module only has imperfect representations of time and counting. Note that the magnitude module is not used for exact, verbal counting, but rather for non-verbal numeric estimation (exact verbal counting can be performed easily by traditional ACT-R).

A key component to non-verbal counting is deciding when to stop. We propose here that the internal temporal module (Taatgen, Van Rijn, & Anderson, 2007) is used. The temporal module tracks time intervals and is quite accurate at short timer scales, becoming progressively less accurate and noisier at longer time scales. The temporal module simply keeps track of how long it has taken since counting began. A rate of counting is calculated based on the (noisy) timer and an updated previous magnitude. Finally, a target amount of time can be determined based on the rate and the target number.

High level description of the magnitude module

There are three components to each model: start, count, finish.

Start The model prepares to begin counting by setting a target-goal (e.g., 17) and preparing to count (e.g., by putting their finger on the counting key). The rate is undefined at this point.

Count The model counts by making a call to the magnitude module for every count it makes. Every count initiates a physical keypress as well. Every count, several quantities are updated.

Rate A current rate is calculated based on the amount of time that has passed since counting began and the successor of the last magnitude.

Magnitude The current magnitude is calculated based on current time and the current rate. Note that because magnitude is based on the model's imperfect sense of time and an imperfect rate, it never has a perfect representation of count. Because the timer is more accurate at short time intervals, it is frequently (but not always) correct at smaller counts. Subitizing is not explicitly modeled and in fact previous researchers have suggested that subitizing is not needed during non-verbal counting (Cordes et al., 2001).

Time-to-stop Time to stop is based on the rate $\times$ target-goal. Because people have different levels of accuracy for non-verbal counting, a mean-scalar (m) and a standard-deviation-scalar (sd) are included in this calculation.

Notice that magnitude ends up having scalar variability. In this account, scalar variability arises because of the imperfect time sense that people have.

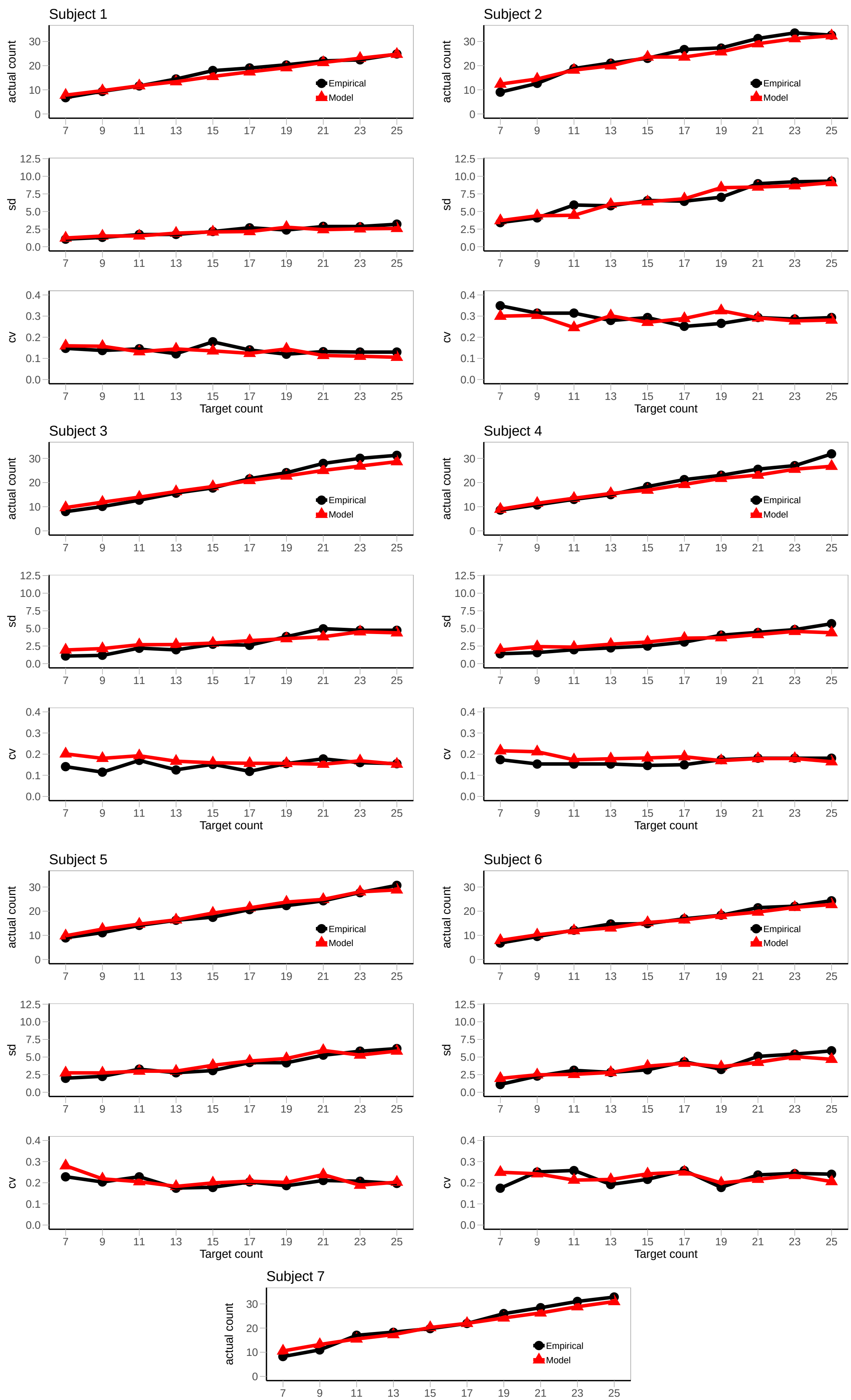
Finish The model finishes counting when the current time is greater than or equal to the computed time-to-stop.

These three components occur in the natural order: Start begins a trial, while Count performs the counting itself, and then Finish ends the trial.

Model Fit

The data was presented in the original Whalen et al. (1999) study as a series of graphs of the seven individuals. A single graph of average performance by participant was not presented (presumably to show that the coefficient of variation was constant across every single participant). The individual data was digitally extracted and averaged into the graph shown in Figure 1. A model was fit to every single participant as well, show in in Figure 2.

Model fits were created by running the model 250 times for both the overall average and each individual. 250 was selected because it provided stability across the entire range of participants and variables. All standard ACT-R parameters were left at their defaults. Two magnitude parameters (m and sd) were fit for each participant and for the average performance. Both parameters stayed within a narrow range (.1 - .7 for m and .3-.5 for sd); changes to these parameters only impacted the strength of the individual fit, not the overall pattern.



Subject 1
Subject 2
Subject 3
Subject 4
Subject 5
Subject 6
Subject 7
actual count
sd
cv
Target count
Empirical
Model

Participant	R^2	RMSD
1	.99	.48
2	.99	7.1
3	.99	3.5
4	.99	2.3
5	.99	4.0
6	.99	1.2
7	.99	4.9
All Participants	.99	3.2

Table 1: Model fit table for counting accuracy (top panels).

Participant	R^2	RMSD
1	.79	.35
2	.87	.70
3	.92	.67
4	.92	.63
5	.90	.53
6	.88	.74
7	.86	.54
All Participants	.94	.42

Table 2: Model fit table for counting standard deviation (middle panels).

R^2 and RMSD fit metrics between the empirical and model data were generated for each individual participant and the average of all participants. Table 1 shows the fit metrics for the count data (how accurate the counting was; top panels). Table 2 shows the fit metrics for the standard deviation data (how the standard deviation increased across target count; middle panels). Table 3 shows the fit metrics for the coefficient of variation data (the relatively constant values across target count; bottom panels). For count data and standard deviation, a high R^2 and a low RMSD shows a good fit. For the coefficient of variation fit metrics, R^2 should be close to 0 because it is a constant, while RMSD should be low.

As can be seen in the fit tables and the graphs, the model fits the data quite well on all three primary variables: count, standard deviation, and coefficient of variation.

Participant	R^2	RMSD
1	.02	.02
2	.04	.04
3	.02	.04
4	.04	.03
5	.04	.02
6	.00	.04
7	.09	.02
All Participants	.06	.02

Table 3: Model fit table for counting coefficient of variation (bottom panels). Note that the R^2 should be close to 0.

General Discussion

We described a process model for non-verbal counting. Our model has several advantages over existing models. First, current models typically rely on an internal representation that is perfect – of magnitude, time, or memory. Consistent with most of people's representations, we believe that none of these are represented perfectly.

The current model does not have a perfect model of time. Previous models use 'clock time' to calculate rates and therefore magnitude. However, there is a great deal of evidence that people's sense of time is quite good for short intervals and becomes worse at longer intervals (Matell & Meck, 2000; Taatgen et al., 2007). Thus, this model uses a cognitively plausible measure of time intervals (Taatgen et al., 2007).

The current model does not have a perfect model of magnitude. Magnitude is represented as a scalar value that increases over time and in the non-verbal counting task we have modeled here it is created directly from the rate of counting. The model suggests that magnitude estimation is inherently imperfect because people do not have a perfect representation of time.

The current model does not have a perfect representation of memory, though it inherits that memory imperfection from ACT-R (Altmann & Trafton, 2002). In the current model, memory is not explicitly used, but certainly if the model needed to store, remember, and retrieve a magnitude the machinery exists to do so.

The current model also solves the sampling problem discussed earlier. Because this model determines when to stop based on time, this model never has a negative or backwards-going magnitude. Nor does this model consistently undercount because of a greater chance of skipping numbers.

The current model can presumably explain non-verbal counting in animals as well. Animals seem to represent magnitudes in the same way that people represent non-verbal magnitudes (Church, 1984; Gibbon et al., 1984; Meck & Church, 1983), and this model would capture the same features (e.g., scalar variability) of animal counting that have been described in the literature (Platt & Johnson, 1971).

It is interesting to note that both magnitude and time sense have similar representations: they both have scalar variability, more accurate at smaller numbers and less accurate at bigger numbers. This remarkable similarity suggests that both time and magnitude are intimately connected. In our model, we connect them directly: people's sense of time is critical to how magnitude estimations occur. Without a sense of time (or if time-sense is being used for something else), the model suggests that magnitude estimation is exceedingly difficult – perhaps so difficult that another strategy would need to be used.

ACT-R is well known for modeling average behavior, and equally well known for not being able to model variability very well. A typical model fit, for example, shows empirical means and model means overlapping. However, these models very rarely adequately model the variability inherent in the empirical data. This model, however, models not only the mean data, but also the variability. This emphasis on modeling the full distribution of behavior is a core strength of our approach here.

We should emphasize that the current model is for non-verbal counting only. Other researchers have studied other forms of numerosity – estimating the number of objects on a screen; explicit counting; approximate counting, and others. Exactly how this model will scale to those other tasks is for future work. Certainly a similar model could presumably capture the observed empirical patterns: examining density and then extrapolating based on how long it took to determine density may be a method to estimate the number of objects on a screen.

In summary, the current model emphasizes non-verbal counting using cognitively plausible – and imperfect – core mechanisms. We modeled one of the best known empirical examples of non-verbal counting (Whalen et al., 1999) and it is the only existing model we know of that captures the full range of non-verbal counting through a high-fidelity process model.

Acknowledgments

This work was supported by the Office of Naval Research to JGT. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies of the U. S. Navy. We thank Sunny Khemlani, Anthony Harrison, Hillary Harner, and Gordon Briggs for their advice and comments on a previous draft.

References

Altmann, E. M., & Trafton, J. G. (2002). An activation-based model of memory for goals. *Cognitive Science*, 39-83.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.

Bell, T. (1937). Complete independence. *Men of mathematics*.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Church, R. M. (1984). Properties of the internal clock a. *Annals of the New York Academy of Sciences*, *423*(1), 566–582.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic bulletin & review*, *8*(4), 698–707.

Dormal, V., Seron, X., & Pesenti, M. (2006). Numerosity-duration interference: A stroop experiment. *Acta psychologica*, *121*(2), 109–124.

Gallistel, C., & Leon, M. (1991). Measuring the subjective magnitude of brain stimulation reward by titration with rate of reward. *Behavioral neuroscience*, *105*(6), 913.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in cognitive sciences*, *4*(2), 59–65.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of sciences*, *423*(1), 52–77.

Klahr, D. (1973). A production system for counting, subitizing and adding. In *Visual information processing* (pp. 527–546). Elsevier.

Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays*, *22*(1), 94–103.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320.

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental psychology: animal behavior processes*, *11*(4), 591.

Meck, W. H., & Williams, C. L. (1997). Characterization of the facilitative effects of perinatal choline supplementation on timing and temporal memory. *Neuroreport*, *8*(13), 2831–2835.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519.

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation*, *2*(4), 386–414.

Taatgen, N. A., Van Rijn, H., & Anderson, J. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review*, *114*(3), 577.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749.

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current directions in psychological science*, *6*(1), 12–16.