Negen, J

 The Numeric Ebbinghaus Effect: Evidence for a Density-Area Mechanism of Numeric Estimation?

http://researchonline.ljmu.ac.uk/id/eprint/21681/

Article

For more information please contact researchonline@ljmu.ac.uk

Empirical Research

Check for updates

The Numeric Ebbinghaus Effect: Evidence for a Density-Area Mechanism of Numeric Estimation?

James Negen [1]

[1] *School of Psychology, Liverpool John Moores University, Liverpool, United Kingdom.*

Abstract

One model of numeric perception is a density-area mechanism: a process that estimates both density and area of an array, then multiplies them to create an estimate of number. One line of evidence that supports this is the surprising numeric Ebbinghaus illusion: smaller context circles lead to greater perceived number than larger context circles, potentially via larger perceived area. This registered report re-tested this effect with a number of simple but potentially important improvements in the method and analysis. Participants were asked to indicate the number of blue dots in arrays that were surrounded by grey context circles of three different sizes. Both experiments confirmed that larger context circles lead to a proportional increase in perceived number. Experiment 1 (N = 50) did so with denser, more texture-like arrays (50-100 dots filling 35% of the area). Experiment 2 (N = 50) did so with sparser, more scatter-like arrays (10-30 dots filling 5% of the area). These findings confirm the existence of the numeric Ebbinghaus effect. This in turn confirms a specific prediction derived from a density-area mechanism and rules out alternatives that begin by stripping away context to non-verbally count discrete entities. No further significant evidence was found to suggest that this depends on the array being particularly dense or texture-like, nor to suggest that anything moderates the impact of increasing perceived area as a direct proportional effect on increasing perceived number. This further builds the case that this kind of numeric perception relies on a density-area mechanism.

Keywords

numeric perception, approximate number system, Ebbinghaus effect, number sense, scalar variability

Adult humans have the capacity to somehow estimate the approximate number of items in a brief visual array without verbal counting (Feigenson et al., 2004). There are detailed, extensive debates over the exact nature of the cognitive mechanism involved but this much is clear: when shown a simple stimulus with a number of dots or rectangles, the average person can at least give a verbal response that covaries with the correct count. There will be noise and there can be biases, but there is still some unique relation present between the verbal response and the correct count (Leibovich & Henik, 2014). Further, this can be accomplished without explicitly counting the items. This means there must be some perceptual/cognitive mechanism, some set of steps for information processing, that begins with a visual array and quickly ends with a verbal number in such a way that some perceptual sensitivity is demonstrated. This capacity, in

addition to being a potential basic aspect of our perception, may also underly important outcomes like STEM education scores (Halberda et al., 2008). It is therefore important to better understand the mechanisms of numeric estimation.

This article explores one category of ways that rapid visual-to-verbal numeric estimation might be accomplished: by creating a density estimate and scaling that by an estimate of the total area of the array. For brevity this can be called a *density-area mechanism.* Figure 1 illustrates the basic principle that allows this to work. This study was designed mainly around improved testing for a surprising prediction derived from a density-area mechanism. It was also designed, whether that prediction is confirmed or not, to create new theoretical boundaries on the cognitive processes that might explain this kind of numeric estimation.

Figure 1

A Simple Illustration of the Mathematical Principle That Makes a Density Area Mechanism Function
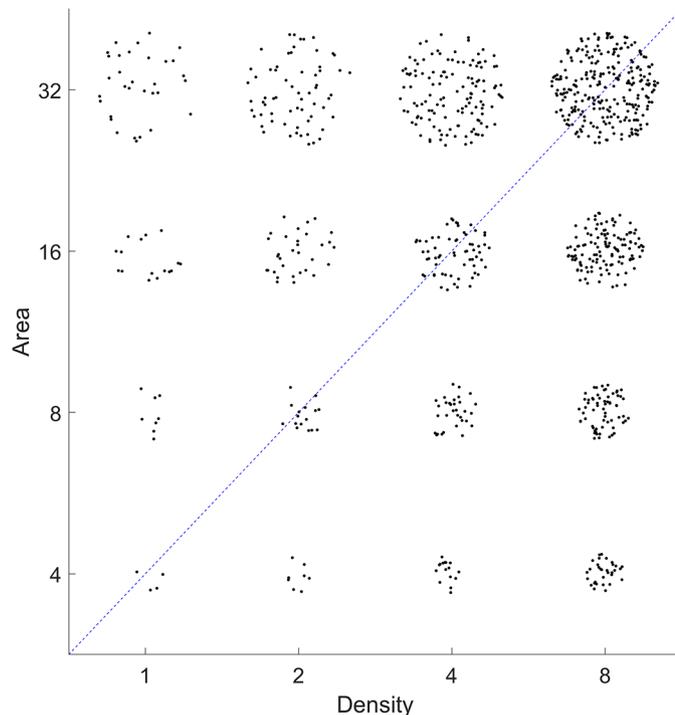
Note. These example arrays are double in density with each unit along the x axis and double in area with each unit along the y axis. Any movement directly upwards or directly rightwards leads to a scaling (in this case doubling) of the number. The number stays the same when moving perpendicular to the blue line. Any two can be compared by projecting perpendicular lines back to the blue line.

The introduction now goes on to (a) provide a general background for context, (b) review a new surprising finding that is predicted by a density-area mechanism, (c) point out a few places where the methods might be varied or improved to see if we can solidify that finding, (d) explore and explain what will be learned if the finding is / is not confirmed, and (e) give an overview of the method and a specific statement of the hypothesis.

General Background

This section will provide a brief overview of the state of research and theory in a few key areas: numeric perception, the Ebbinghaus illusion, and the perception of density.

Numeric Perception

Numeric perception/cognition can be broadly divided into three sections: symbolic, non-symbolic small, and non-symbolic large (Carey, 2009; Feigenson et al., 2004). Symbolic representations of number include things like the written

number 6 or the spoken word "eleven". These are arguably special in their capacity to be durable, precise, compact in memory, and unbounded in their extent (Carey, 2009). Non-symbolic small numbers include things like arrays of just 1-4 dots which can be tracked with extremely high accuracy (Feigenson et al., 2004; Revkin et al., 2008). Non-symbolic large number then includes the rest – every way that number is estimated without counting, without a symbolic representation of the number, and without relying on the specialized mechanisms that can be employed for very small set sizes. While this last category does include multiple modalities (Feigenson et al., 2004), this study focuses on the perception of non-symbolic large number in visual arrays.

Theories for how non-symbolic large number is perceived have broadly fallen into three categories (Leibovich et al., 2017). Degenerate theories suggest that number is not perceived, but rather that its supposed perception is an artifact of poor experimental control. This has been a much more serious consideration for very young participants than adults. Discrete models suggest that a visual array is first processed into individual units and then somehow non-verbally counted (Cordes et al., 2001; Meck & Church, 1983). Continuous models suggest that some continuous aspects of the array – things like area, dot size, brightness, and so on – are sensed first and then used to create a numeric estimate. The contrast between discrete versus continuous also typically maps to domain-specific versus domain-general: a discrete mechanism begins by stripping away non-numeric information whereas a continuous mechanism takes non-numeric aspects as its input. In principle there could be continuous domain-specific theories but these seem to attract less interest.

Within the range of continuous domain-general theories, there are then a variety of more specific proposals. One prominent kind of these is the use of an area estimate to scale a density estimate (Dakin et al., 2011). Such ideas have seen a recent revival (Leibovich et al., 2017), though it is worth noting for general background that some approaches fall within the continuous domain-general category without necessarily being a density-area mechanism (e.g. Stoianov & Zorzi, 2012).

It is also important to note that some recent work suggests a difference between the way more dense (texture-like) displays are estimated versus the way more sparse (scatter-like) displays are estimated. For example, it has been argued that the noise in sparse arrays follows Weber's law whereas the noise in dense arrays follows a square root discrimination law (Anobile et al., 2014). It has also been argued that sensitivity to number is far greater than sensitivity to density or area in sparse arrays (Cicchini et al., 2016), arguing in particular against a density-area mechanism for sparse arrays. The two experiments therefore focus on more dense (Experiment 1) and more sparse (Experiment 2) arrays to see if the numeric Ebbinghaus effect is present in each.

Within this context, the present study seeks to cut directly to mechanisms. The present study sets aside the popular questions of "innate", "holistic", "pure", or "basic" (Leibovich et al., 2017); it seems unclear what evidence would resolve these. Rather, it takes the approach of focusing on surprising hypotheses derived from specific mechanistic theories. It is left to the reader to decide which (if any) of these controversial adjectives apply to the described mechanism.

Ebbinghaus Illusion

The Ebbinghaus illusion can be harnessed to reliably create an illusory change in area. The simplest version of this illusion uses two central dots of the same size that are each surrounded by context circles. The context circles around one central dot are larger than the ones around the other central dot. This creates the perception that the central dot surrounded by smaller circles is larger. While this can be partially explained by a contrast effect (between sizes of the central dot versus context circles), modern research has discovered that it also has to do with more than just relative size (Roberts et al., 2005). It is actually possible for context circles that are smaller than the central dot to make the central dot look smaller if they are positioned correctly (ibid). In other words, the contrast between the apparent size of a no-context dot versus a surrounded dot is difficult to predict. However, when comparing one display with context circles against another, we can still predict higher perceived area if one display has smaller circles, more circles, and less distance between the context circles to the central object, especially if both sets of context circles make an almost-complete ring around central object (ibid). This is the approach taken here: context circle size, number, and distance all covary and the context circles always form a near-complete ring in all the stimuli used here. This means that a stimulus with smaller context circles should always be perceived as larger.

PsychOpen GOLD

Density Perception

Recent theory also posits that density (or more likely, some close approximation to it) is an output of low-level visual perception (Durgin, 2008). This can be demonstrated by a brief fatigue after-effect (see their Figure 2). Much like staring at a bright green patch will create a red after-image, staring at a highly dense patch of dots creates a low density after-image. The calculations involved are something like kurtosis (Dakin et al., 2011; Durgin, 2008) and could likely be handled by cells as early as V1 in visual cortex (ibid). This suggests, for our purposes here, that density is a plausible input into the mechanism that estimates number.

With that said, the use of density in the theoretical sketch here is an attempt to understand the principles of the algorithm and is unlikely to have a perfect reflection in the implementation. As an analogy, the concept of multiplication in mathematics is very rich, covering things likes irrational numbers and alternative (non-Peano) systems, and it also has infinite precision. The implementation of multiplication in a typical computer program does not have the same richness, usually failing to properly preserve irrationality and only working for typical (Peano-based) systems, and it always necessarily has finite precision. These schisms don't defeat the way that these implementations can still be usefully understood as multiplication. Much of our everyday life depends on software where nobody has particularly thought through the difference. In much the same way, it is likely that density is a principle that has a useful approximation in the actual brain implementation.

To clarify, it might be helpful to look at an example of recent work that examines brain implementation of numeric perception. Recent work (Paul et al., 2022) shows that aggregate Fourier power increases monotonically with number in a given local stimulus area with little impact of object size, shape, or spacing. They further show that V1 responses reflect this closely with increasing cortex response amplitude. Importantly, this sensitivity to aggregate Fourier power is retinotopic. In short, it suggests that local V1 areas track the local frequency energy and report it to later numerosity calculations (though with a gate for bounded object presence). This can be understood as implementing an approximate retinotopic map of local number, also known as density, in a way that ignores or filters some common confounds – even though the implementation also has a perfectly valid description in terms of spatial frequencies and energies that never exactly mentions density.

The Numeric Ebbinghaus Illusion

This section lays out a key previous finding that the present study will build upon heavily. A recent paper has suggested a surprising new effect that is consistent with the theory of a density-area mechanism: the Ebbinghaus illusion, where a smaller set of context circles is used to make a central array look larger, will also influence the perceived number of items in the same direction (Picon et al., 2019) (Figure 2).

Figure 2

Example of the Numeric Ebbinghaus Illusion With Stimuli From the Present Study
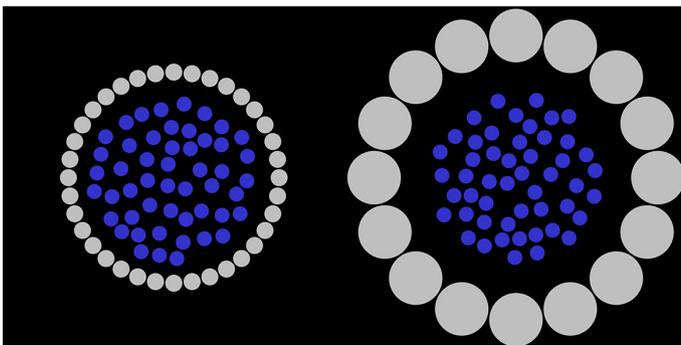
Note. Previous research suggests that participants will usually give a higher estimate for the number of blue dots in the left array (Picon et al., 2019) even though they each contain 50 blue dots.

This is obviously consistent with a theory where perceived number is found by multiplying perceived area by perceived density; increase the perceived area and the perceived number will also increase. Critically, this key effect occurs regardless of whether the response is a choice or an estimate (ibid). Because of this, it cannot be easily explained by response bias. This finding not only confirms an interesting and surprising prediction from a density-area mechanism, but also explicitly rules out any theorized mechanism that begins by stripping away context and continuous features (e.g. Cordes et al., 2001). This makes it an important finding for our understanding of numeric perception – which, in turn, justifies why we must examine the evidence for this finding critically and improve upon it where needed.

Potential Issues With the Evidence

This section explains how and why there is scope for a more solid base of evidence for the numeric Ebbinghaus effect. To begin, there is clear scope to test a more specific prediction. In particular, the previous study (Picon et al., 2019) tests the broad idea that the Ebbinghaus illusion will influence numeric estimates. If number is estimated by multiplying area and density, then we can predict that a change in perceived area should result in a *proportional* change in perceived number. For example, if a 1.0 units^2 area is perceived as 1.25 units^2, then the perceived number of items should be biased upwards by 25% (rather than a constant or a function of 1.25^2 and so on). This can be tested explicitly by extracting separate parameters for a constant change, proportional change, and squared change to check that the proportional one retains the loading.

Closely related to the point above, estimating a separate constant effect is a vital control against a serious potential confound. This is because smaller context circles require more context circles to create a similarly complete surround. In other words, look at Figure 2: if the left array appears more numerous, this could either be (1) because of an Ebbinghaus effect influencing area perception and thus numeric perception or (2) because it actually has 20 more circles in it (they just happen to be grey). In the present method, the smaller, medium, and larger context circles were always presented in sets of 36, 25, and 16 context circles. Participants were instructed to estimate the number of blue dots with the word "blue" in bold. If they either ignore this or find it impossible, then we would specifically expect a constant increase of 36 – 16 = 20 more dots when comparing the smaller versus larger context circles. Again restated, we needed to separately estimate a constant and proportional effect to be sure that previous results were not just reflecting this specific confound (constant effect) but rather the type of effect one would expect from a density-area mechanism (proportional effect). Until this is established, it is not clear that previous results reflect any effect on the perceived numerosity of the central array.

It would also be helpful to use a full model that deals explicitly with scalar variability, round numbers, and range limits. Previous work (Picon et al., 2019) dealt with scalar variability to some extent by using outcome variables that were scaled to the true number, capturing proportional error. However, this is slightly off-target; scalar variability suggests that the standard deviation is proportional to the *mean response* rather than the true target number (Cordes et al., 2001). This detail is likely negligible but still good to control in the context of large biases. Further, it is well-known that participants are unlikely to give responses like 81 dots, 82 dots, 83 dots, or 84 dots (e.g. Jansen & Pollmann, 2001). Almost all will almost always say either 80 or 85. This means that a response of "85" does not really mean that they perceived exactly 85; rather, it indicates that they think it is between 82.5 and 87.5. While this again is likely negligible, there is no particular reason not to deal with it explicitly and fully. Along similar lines, participants are likely to infer (either correctly or not) that there is some range limit on true answers (e.g. all answers will be between 50 and 100). This would make them unlikely to give responses outside the range, distorting the response distribution away from the distribution of their perceptions. All three of these issues were managed by (a) giving people a choice of pre-determined round responses that match the true range (Figure 3) and (b) applying a simple model for parameter estimation.

It would also be helpful to deal with calibration and feedback in a more explicit manner. The previous paper (Picon et al., 2019) does not particularly state if feedback was or was not given. The pattern of systematic overestimation suggests that it was not. The analysis of this kind of task will typically assume that some stable calibration exists between non-verbal perception of number and verbal outputs for number. It might be a major issue if this calibration is not set firmly and maintained; the same verbal response might indicate different underlying perceptions at different points in the experiment. This could damage statistical power, lead to underestimation of effects, or lead to these calibration

changes being mistaken for effects of interest. This concern is not purely hypothetical: extensive miscalibration has been documented with untrained adult participants in previous studies (Sullivan & Barner, 2013). It therefore seems prudent to provide enough feedback to become calibrated and maintain that calibration. On the other hand, feedback on every trial would enable and encourage people to explicitly compensate for the biases of interest. The best method will therefore provide feedback strategically to maintain calibration but not correct the biases of interest.

Figure 3

Screenshot From the Task Demonstrating the Response Mechanism
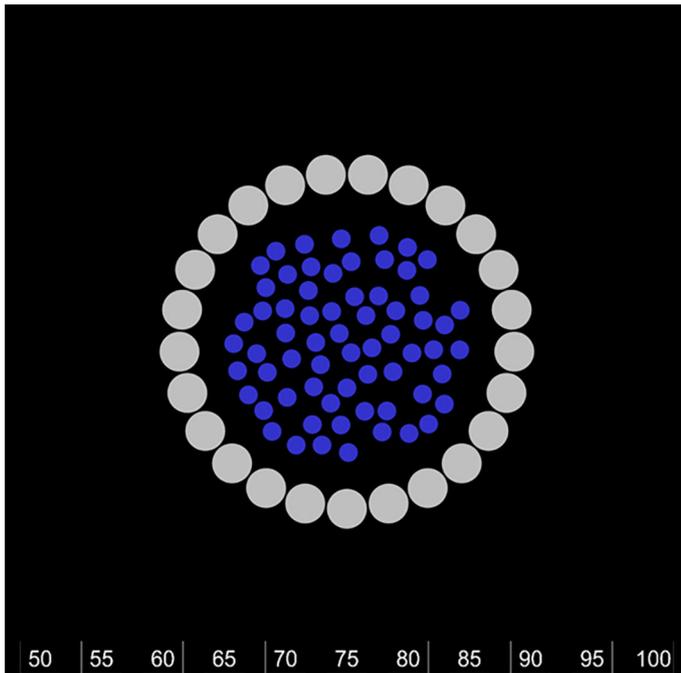
Note. Participants click/tap on one of the pre-selected round numbers.

Finally, it would also be helpful to measure the non-numeric Ebbinghaus effect with the same context circles for comparison. One could imagine either a moderated or unmoderated version of a density-area mechanism – one that tempers the scaling effect and uses other continuous features versus one where increases in perceived area are reflected one-to-one as proportional increases in perceived number. This can be resolved by asking participants for a simple adjustment to create a perceptual match, indicating the size of the non-numeric Ebbinghaus effect as a reference. This context extends the results here beyond a simple binary decision, extending the work here beyond previous study.

    To summarize, we can build up a better base of evidence by (i) testing the proportional prediction as separate from a constant confound, (ii) dealing with scalar variability in the response model, (iii) dealing with round numbers in the response mechanism and model, (iv) dealing range limits in the response mechanism and model, (v) providing feedback that is sufficient for continuous calibration but not for correction of the biases of interest, and (vi) also measuring the non-numeric Ebbinghaus effect with the same context circles.

New Limits on Theory

This section lays out exactly how the study's results could affect theory. The possible results effectively fall along a scale: the proportional numeric Ebbinghaus effect is (a) not significantly above zero and significantly less than the non-numeric effect; (b) significantly above zero but significantly below the non-numeric effect; (c) significantly above zero and not significantly less than the non-numeric effect. (If neither comparison is significant, the study will be

withdrawn as failing to meet outcome-neutral criteria.) These three possible results then lead on to three possible interpretations.

Case C, where the numeric effect is above zero and not significantly below the non-numeric effect, is consistent with a density-area mechanism and perhaps even an unmoderated one. The changes to the context circles (size, number, and distance) will change the perceived area. If the estimate of area is a key scaling factor then it stands to reason that density must also be involved since a density factor is required to transform area into number. This helps build a case for some type of density-area mechanism, at least for this type of numeric perception (adults, rapid, larger sets, visual-to-verbal estimation). It further fails to give any particular reason to include any attenuating mechanism; it is consistent with the simplest kind of density-area mechanism. As with all three outcomes, confidence intervals around the effects will create reference points for any future work that wants to posit specific effects of the present manipulation. Any future computational model would need to respond to the stimulus changes within the calculated confidence interval.

Case B, where the numeric effect is above zero but below the non-numeric effect, is consistent with a kind of moderated density-area mechanism. Since the effect is above zero, the involvement of area (and thus density) is consistent with this case by the same logic as above. However, this would indicate that a particular increase in perceived area is not reflected fully in perceived number. There could be various explanations – perhaps there is a domain-specific estimate of area that is also used; perhaps other covariates attenuate the effect; perhaps some mechanism attempts to find illusory size changes and partially compensates for them. In any case, this will help build a case for a continuous mechanism that uses area as a moderated scaling input.

Case A, where the numeric effect is not above zero but is below the numeric effect, would speak towards the idea that domain-general area perception is not a direct input into numeric perception – or at least, a minor one. This would be consistent with the idea that the previously reported effect is due to methodological issues rather than cognitive mechanisms. Such a failure would leave continuous domain-general theories in the position of explaining how number is estimated while either (a) not using area as a scaling input or (b) using it only to the upper extent of the calculated confidence interval. If this could not be achieved, that would in turn favour discrete and/or domain-specific mechanisms.

The reader should be aware that cases B and C are consistent with a density-area mechanism but could potentially have other unspecified explanations. In theory, we know that the Ebbinghaus illusion causes a change in perceived area. However, it also affects various low-level visual statistics that might feed into various mechanisms not considered here. It is always possible that some future theory will explain the numeric Ebbinghaus effect without a density-area mechanism. In case B or C, an explanation of this will be present in the Discussion.

Summary Method and Hypothesis

The basic method was to show participants arrays of blue dots and ask them to estimate the number. The key manipulation was that the blue dots were surrounded by context circles of three different sizes on different trials. Feedback was given on the trials where the context circles were the middle size, creating a calibration signal but not giving the feedback needed to compensate for the context circle effect. To discourage degenerate strategies based on individual dot size, the total area of the array was also varied. The primary hypothesis was that the proportional numeric Ebbinghaus effect would be above zero and not significantly different from the non-numeric effect with the same context circles.

Experiment 1

Method

The full method, analysis code, and all data are on the Open Science Framework (Negen, 2023b). This includes a pre-registration (Negen, 2023a) which occurred after the preliminary data set (which was pilot work to see if the method was plausible) but before the two final data sets (which are the basis of the conclusions here).

Participants

Preliminary results were gathered with 10 participants (ages 27 to 76, mean 38, standard deviation 14; 6 male, 3 female, 1 gender fluid). None were excluded for showing $r < 0.3$ between target and response. Participants were recruited via Prolific (https://www.prolific.co/). They were screened for fluency in English and normal or corrected-to-normal vision. Ethics approval has been granted by the Liverpool John Moores University Research Ethics Committee (22/PSY/027 "Numeric Estimation").

The full study used an additional sample of 50 participants per experiment via the same recruitment method and with the same screens. The sample size was primarily based on the large effect sizes found in previous research (Picon et al., 2019). Looking at their Figure 4, they found a between-condition difference of at least 0.325 with a standard error of at most 0.08 with 18 participants, leading to a Cohen's d of at least $0.325/(.08*18^{1/2}) = 0.95$. For the same effect size, the present study would have over 99.9% power. It would have to come down to $d = 0.35$ to break below 80% power. This should be more than adequate to detect the effect of interest if it exists, especially since the custom model below is designed to lead to more accurate parameter estimation.

Apparatus

Participants participated online using their own computers or tablets. This was enabled by Pavlovia (https://pavlovia.org/). On the screen there could be three things: the stimulus, the response buttons, and the feedback indicator. The stimulus was presented in a square with a side length that equals 90% of the larger screen dimension. Below it were the response buttons: 50 through 100 in steps of 5. This was consistent with the responses that most participants give as free responses anyway; few participants will estimate a non-round number like 83 dots. The buttons were in a line and separated slightly to hopefully reduce accidental presses. The feedback indicator was a simple green triangle that could appear over the correct response. While participants may vary in terms of distance to the screen, screen size, and so on, the analyses below all depended on within-subjects criteria that should not be affected.

Stimuli

There were a total of 99 stimuli generated for the first experiment. This was a full factorial design for true N (i.e. items in the array; 11 levels: 50 through 100 in steps of 5), a scaling factor for the total area of the stimulus (3 levels: 100%, 75%, and 50% scaling in terms of area), and the size of context circles (3 levels: 2.5%, 4%, and 7.8% of the stimulus width as their radius). Figure 4 below illustrates the effects of varying these three dimensions. The size of each blue dot was constant within each stimulus but varied between stimuli. The blue dots always took up 35% of the total area for the array. Blue dots were placed randomly such that they did not overlap but fell inside the total area of the array. The blue dots were specifically RGB values of 0.2, 0.2, and 0.8. Context circles were 75% grey. The background was pure black. There were 36, 25, or 16 context circles depending on their size. The context circles were placed with one radius of buffer between their inner edge and the outer edge of the area for the blue dots. All stimuli were initially rendered at 4000 by 4000. They were then scaled down as appropriate depending on the area scaling factor. Black pixels were added around to again bring it back to 4000 by 4000. This was then again sized down to 1000 by 1000. A full copy of all stimuli can be found on OSF with the rest of the method (see Negen, 2023b).

Procedure

First, an instruction slide was given (Figure 5). Second, there were 11 warmup trials. These used each N but with the middle size for context circles and the 100% scaling factor. Third, the main testing trials were run. These involved all 99 stimuli in random order, once each. Each trial went through the same basic steps: A fixation cross appeared for 1.5 seconds. The stimulus was shown for 500ms. The response buttons activated (they were visible before this but did not do anything if pressed). After a response was entered by pressing one of the buttons, if the trial's context circles were the middle size, a green feedback triangle appeared over the correct answer for 2 seconds. This meant that participants were presented a stream of calibration information but only with the middle context circle size.
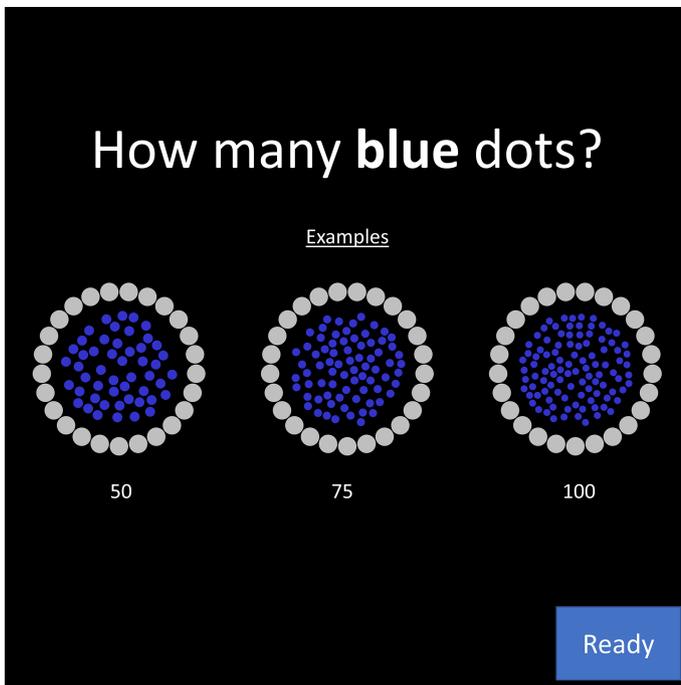
Figure 4

Four Example Stimuli to Demonstrate Variation Along the Three Dimensions
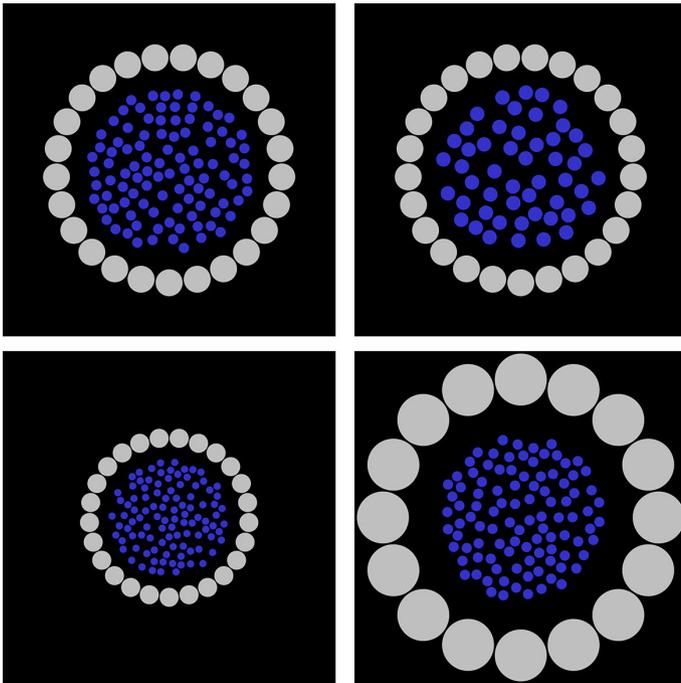
Note. The top left has 100 dots, an area scaling factor of 100%, and the middle context circle size. The top right instead has 50 dots. The bottom left instead has an area scaling factor of 50%. The bottom right instead has the larger context circle size.

Figure 5

Instructions Given to Participants

After the main experiment was an adjustment trial intended to measure the non-numeric Ebbinghaus effect. Figure 6 shows this interface. The smaller and larger context circles were displayed around large blue dots. There were plus and minus buttons to adjust the size of the left blue dot, the one in the smaller context circles. There was also a check mark to end the trial. Finally, there were instructions across the top: "One last thing! Please use the + and - to adjust these until the blue dots appear the same size." The two dots were offset vertically so that they are harder to judge by imagining horizontal lines.

Figure 6

Interface for the Final Adjustment Trial
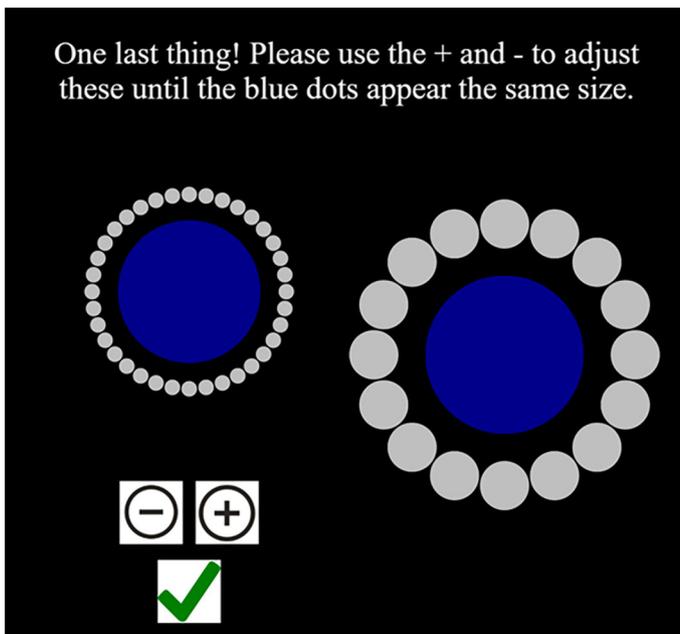
Planned Analysis

The basic method of the analysis plan was to (a) extract parameters from each participant separately with a model and (b) to test these parameters against the hypotheses. This needed to be done while also excluding participants who may not have understood the task properly. This was programmed in MATLAB (MathWorks, 2021) so that it could be completed without any further choices being required (see DAPipeline on OSF for exact code).

Extracting the relevant parameters was done with a modelling approach via maximum likelihood. This was chosen because it can deal with three factors that make typical linear regression unsuitable: (1) we expect the standard deviation of response residuals to not be constant, but rather a linear function of mean response (Cordes et al., 2001); (2) a response of "65" really means any perceived quantity more than 62.5 and less than 67.5; and (3) a response of "50" really means any perceived quantity below 52.5, with a mirroring concern for "100". To accomplish this, a probability density function (PDF; a function taking in the data and parameters, then outputting a likelihood) was programmed in MATLAB. A further MATLAB routine then found the parameters that maximize the PDF's output for each participant.

The PDF (which is effectively the model) had ten parameters for each participant. These were beta values for N, total area, two for constant context circle size, two for proportional context circle size, and two for squared context circle size, plus an intercept and a coefficient of variance (CoV). For this, we define I_C^+ to be 1 when the context circle size is on the largest setting and I_C^- to be 1 when it is on the smallest setting (zero otherwise). The expected mean for each trial μ was calculated as

$$\mu = Y + N\beta_N + A\beta_A + I_C^+\beta_C^+ + I_C^-\beta_C^- + NI_C^+\beta_{NC}^+ + NI_C^-\beta_{NC}^- + N^2I_C^+\beta_{N^2C}^+ + N^2I_C^-\beta_{N^2C}^- \quad (1)$$

where Y was the intercept, N was the true number of dots, and A was the total area. The standard deviation for each trial σ was then calculated as

$$\sigma = \mu \times \mathrm{CoV} \tag{2}$$

The final probability of each possible response was then calculated as

$$P'(R|\mu,\sigma) = \begin{cases} \Phi(52.5|\mu,\sigma) & if\ R = 50 \\ \Phi\ (R + 2.5|\mu,\sigma) - \Phi(R - 2.5|\mu,\sigma) & if\ 55 \leq R \leq 95 \\ 1 - \Phi(97.5|\mu,\sigma) & if\ R = 100 \end{cases} \tag{3}$$

$$P(R|\mu,\sigma) = P'(R|\mu,\sigma) \times .98 + \ .02/11 \tag{4}$$

where R was the chosen response, Φ was the normal cumulative density parameterized with a mean and standard deviation, and the .02/11 term reflected an assumed 2% chance that attention lapses and the participant simply guesses. This is likely much easier to understand with a visual example. Figure 7 shows the response probabilities if a trial has $\mu = 85$ and $\sigma = 10$. In the upper panel, the normal bell curve with a mean of 85 and standard deviation of 10 is shown. It is then cut into the sections that would lead to responses of 50, 55, 60, ..., and 100. The lower panel then shows the final (discrete) probability of each response. These correspond to the total area for the matching section above. This is what Equations 3 and 4 are doing: working through how the mean and standard deviation predict the probability of the different discrete responses.

The final processing step was to calculate the proportional numeric Ebbinghaus effect, $\left((1 + \beta_{NC}^-)/(1 + \beta_{NC}^+)\right) - 1$, and non-numeric Ebbinghaus effect, $(A_1/A_2) - 1$, where A_1 is the area in the larger context circles on the final adjustment trial. For example, a value of 0.5 for the numeric effect would indicate that 50 dots appear to be 50% more when surrounded by the smaller circles than the larger circles – perhaps one appearing as 60 and the other as 40. The non-numeric effect was defined in a comparable way: a value of 0.5 would indicate that the ratio of the two dots areas is 150% at the point of subjective equality.

Effects were then tested against the hypotheses. Participants were excluded if either (i) their initial correlation between N and response was under 0.3, indicating poor understanding/attention to the task (e.g. a person who just clicks responses at random for the payment); or (ii) their proportional numeric Ebbinghaus effect was more than 2.5 standard deviations from the sample mean. There was then a one-tailed t-test to see if the proportional numeric Ebbinghaus effect was above zero. There was also a two-tailed paired t-test to see if the numeric versus non-numeric Ebbinghaus were matching. 95% confidence intervals were reported around both tested values (i.e. the numeric effect and the difference) in each relevant results section. The entire experiment would have been withdrawn if neither comparison was significant (this would indicate either that the stimuli are not arranged to make the non-numeric effect as strong as desired or that estimation is much less precise than pilot results would suggest). Since this was not the case, results were interpreted on the A/B/C scale described in the Introduction (see *New Limits on Theory*).

An additional effect was then calculated for context. This was the correlation between the numeric and non-numeric Ebbinghaus effects. This was given a confidence interval and a Bayes factor from Jamovi. As this has potentially lower power than the t-tests described above, it does not become part of the interpretation logic.

To reiterate: MATLAB used the probability density function and the data to find the maximum likelihood estimate for each participant; exclusions were performed automatically by MATLAB; the final results depended on a one-tailed t-test and a two-tailed t-test that compared the proportional numeric Ebbinghaus effect against zero and the non-numeric effect; a correlation between numeric and non-numeric effects was described for context. All of the code for this has been publicly available on OSF dating before the final data collection. The code required no further interventions or decisions (just the new data dropped into the appropriate folder).

Figure 7

An Illustrative Example of How the Model Assigns Probabilities to Each Response
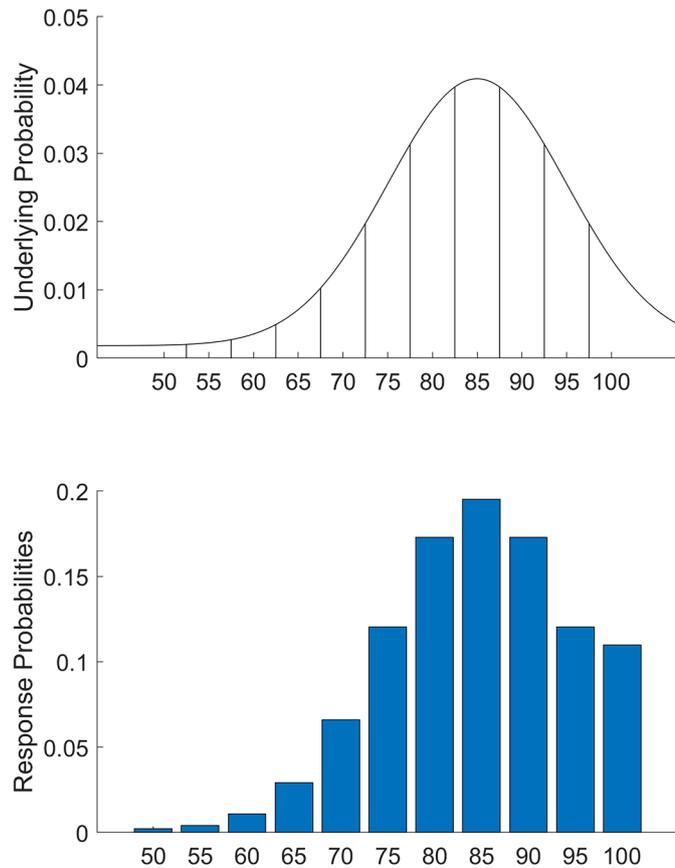
Note. Here, the mean perceived quantity is 85 and the standard deviation is 10.

Preliminary Results

The preliminary results, if they were final results, would be interpreted under case C (Figure 8). The proportional numeric Ebbinghaus effect was significantly above zero, $t(9) = 2.87$, $p = .009$, $d = 0.91$ with a mean of 22%. The difference between the numeric and non-numeric effects was not significant, $t(9) = 1.30$, $p = .226$, $d = 0.41$ with a mean of 10%. This was consistent with the idea that perceived area is a key scaling input into the perception of number, at least in tasks of this type (adults, rapid, larger sets, visual-to-verbal estimation). It also failed to give any reason to posit further mechanisms that might attenuate this effect; it is most directly explained by suggesting an unmoderated density-area mechanism. The correlation between the two effects was estimated at $r = 0.25$, $BF_{10} = 0.48$, 95% CI from -0.39 to 0.69.

In addition, the preliminary results showed three features that were encouraging regarding the capacity of the method to investigate these effects. First, the beta value for N had a mean of 1.02 and standard deviation of 0.07. This was close to the 1.0 value that a perfectly calibrated observer would achieve. Second, the coefficients of variance were all in the expected range, the highest being 0.19. This means that responses had a high signal to noise ratio which should lead to good statistical power. Third, the distribution of responses as a function of N was largely as we would expect: mainly clustered on the diagonal (Figure 8). All three of these observations pointed towards the task being sensible, well-understood, and capable of detecting the relevant trends.
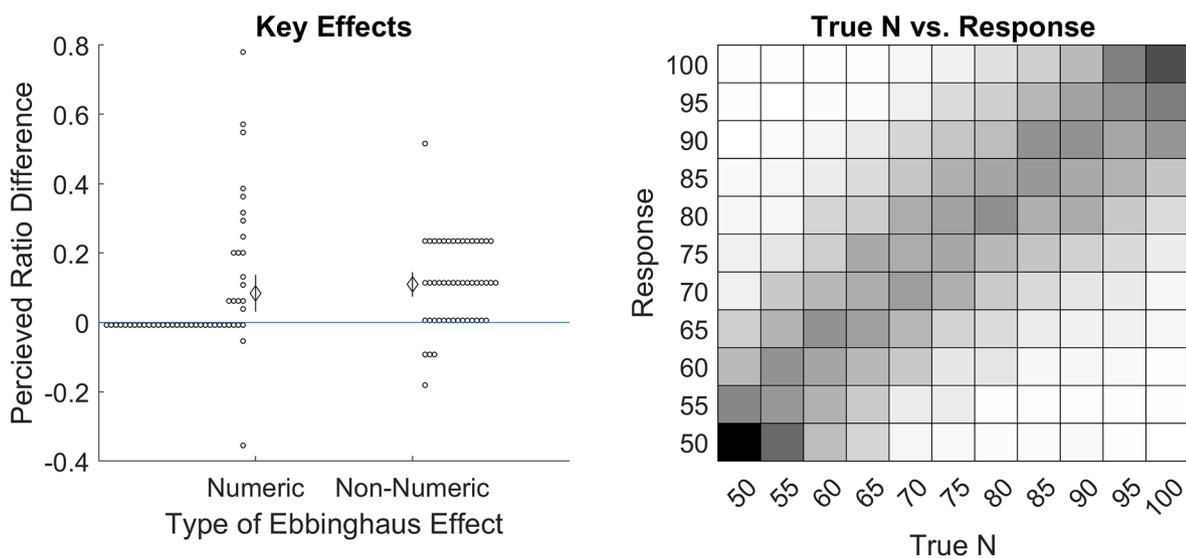
Figure 8

Preliminary Results
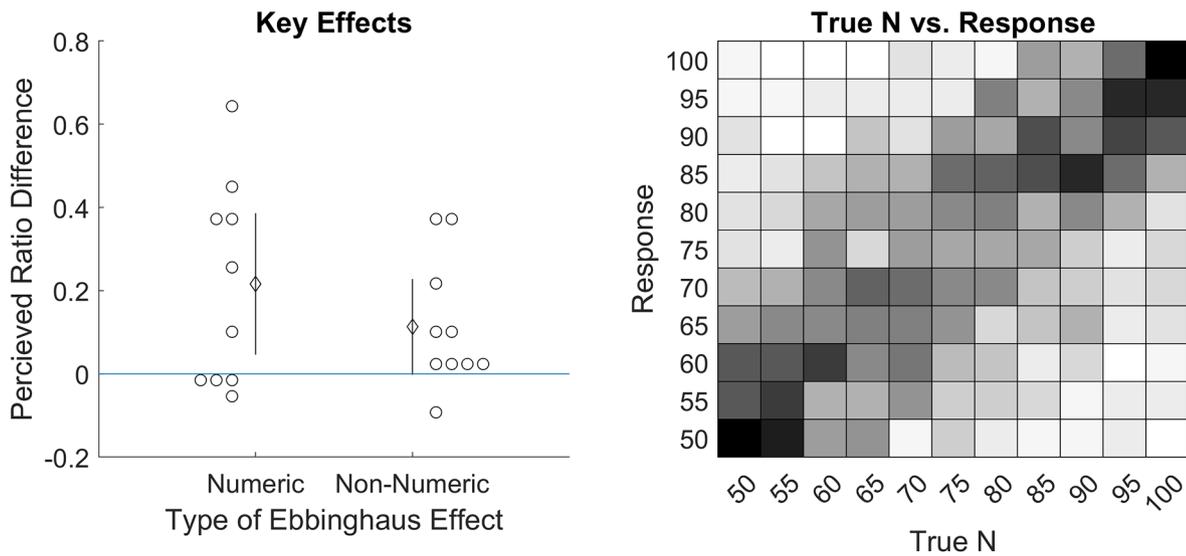
Note. The numeric Ebbinghaus effect was significantly above zero on average but not significantly different from the non-numeric effect (left). Responses mainly fell along the diagonal containing correct responses (right).

Final Results

The final results also fell into case C (Figure 9) with a significant numeric effect and no significant difference to the non-numeric effect. There were 54 initial participants (ages 20 to 80 years, mean 37, *SD* of 13; 27 male, 20 female, and 7 with no response).

Figure 9

Final Results for Experiment 1 (Denser Stimuli)

Note. The proportional numeric Ebbinghaus effect was again significantly above zero on average but not significantly different from the non-numeric effect (left). Responses again mainly fell along the diagonal containing correct responses (right).

One participant was excluded for failing to show a correlation between true N and response above 0.3. Another three were excluded as outliers based on their recorded proportional numeric Ebbinghaus effects (z = -3.21, 2.72, and 4.07). For the remaining 50 participants, the proportional numeric Ebbinghaus effect was significantly above zero, $t(49)$ = 3.19, p = .001, d = 0.45 with a mean of 8% (95% CI from 3% to 14%). The difference between the numeric and non-numeric effects was not significant, $t(49)$ = -0.78, p = .437, d = -0.11 with a mean of -3% (95% CI from -9% to 4%). This was again consistent with the idea that perceived area is a key scaling input into the perception of number, at least in tasks of this type (adults, rapid, larger sets, visual-to-verbal estimation). It also failed to give any reason to posit further mechanisms that might attenuate this effect; it is most directly explained by suggesting an unmoderated density-area mechanism. The correlation between the two effects was estimated at r = -.051, BF_{01} = 5.34, 95% CI from -0.316 to 0.224. This provided moderate evidence that these two effects are not correlated across individuals.
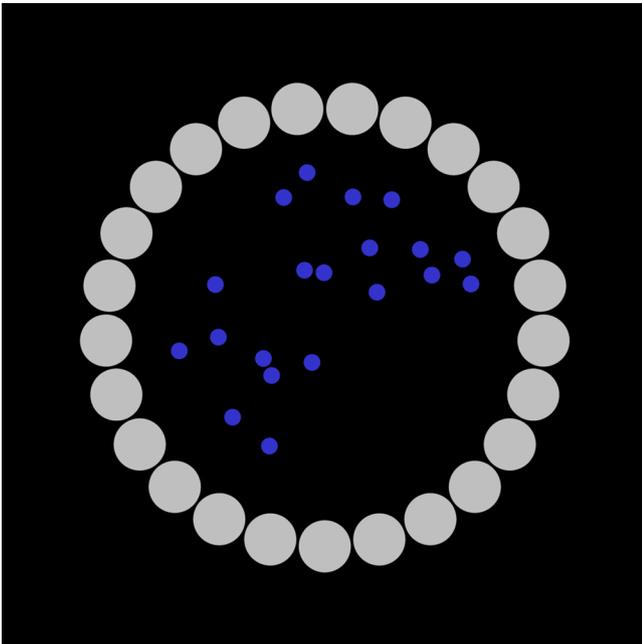
Experiment 2

Method

In Experiment 1, the stimuli used create a relatively high density. Some research suggests that this is a fundamentally different mechanism from displays with much lower density (Anobile et al., 2014). Experiment 2 was very similar but with the following alterations to test lower densities:

1. The number of dots in the stimuli was 10 to 30 (down from 50 to 100), still in increments of 5.
2. The proportion of the overall stimulus area taken up by the blue dots was 5% (down from 35%). Together with #1, this created much more sparse and irregular stimuli (Figure 10).
3. Every stimulus parameter combination was repeated i.e. there were 5 (N) x 3 (area) x 3 (context circle size) x 2 (repeat) = 90 stimuli.
4. The response options and analysis model were adjusted to the new range.
   The new stimuli are available on OSF (Experiment 2 Stimuli).

Figure 10

An Example of the Sparser Stimuli

Note. This one contains 20 dots, the largest total area, and the medium context circles.

This brought up a number of details which are recorded here. Experiment 2 was interpreted independently on the same A to C scale as Experiment 1. Before the final data were collected, it was decided that if one experiment were withdrawn (i.e. no significant difference for the proportional numeric Ebbinghaus effect against either zero or the non-numeric effect) but not the other, the report would be completed with the withdrawn experiment labelled as inconclusive. If both pass the withdrawal criteria and a different case were found, the proportional numeric Ebbinghaus effects for the two would be tested with a *t*-test. Power was not reconsidered to find a difference between experiments as (a) that was not the primary concern here and (b) it is very hard to begin that analysis with a meaningful effect size estimate. Instead, if no significant difference were found between the experiments, the conclusions would include a warning that this should only be interpreted in light of this design decision.

Results

The final results with the sparser stimuli also fell into case C (Figure 11) with a significant numeric effect and no significant difference to the non-numeric effect. There were 51 initial participants (ages 20 to 76 years, mean 36, *SD* of 12; 26 male, 21 female, and 4 with no response). None were excluded for their correlation between true *N* and response. One was excluded as an outlier based on their recorded proportional numeric Ebbinghaus effect (z = 4.60). For the remaining 50 participants, the proportional numeric Ebbinghaus effect was significantly above zero, $t(49)$ = 1.78, p = .041, d = 0.25, with a mean of 6% (95% CI from -1% to 12%). The difference between the numeric and non-numeric effects was not significant $t(49)$ = -1.86, p = .068, d = -0.26, mean of -7% (95% CI from -15% to 1%). This was again consistent with the idea that perceived area is a key scaling input into the perception of number, at least in tasks of this type (adults, rapid, larger sets, visual-to-verbal estimation). It also failed to give any reason to posit further mechanisms that might attenuate this effect; it is most directly explained by suggesting an unmoderated density-area mechanism. The correlation between the two effects was estimated at r = -.054, BF_{01} = 5.31, 95% CI from -0.318 to 0.222. This again provided moderate evidence that these two effects are not correlated across individuals.

Further, the results from the two experiments showed no particular statistical difference. Both resulted in Case C outcomes. The proportional numeric Ebbinghaus effects were not significantly different, $t(98)$ = 0.65, p = .519, d = 0.13 with a mean difference of 2.69% (95% CI from -5.56% to 10.9%). While the second experiment returned a *p*-value nearer the .05 border, there was no clear evidence that this reflects anything more than simple sampling noise.

Figure 11

Results for Experiment 2 (Sparser Stimuli)
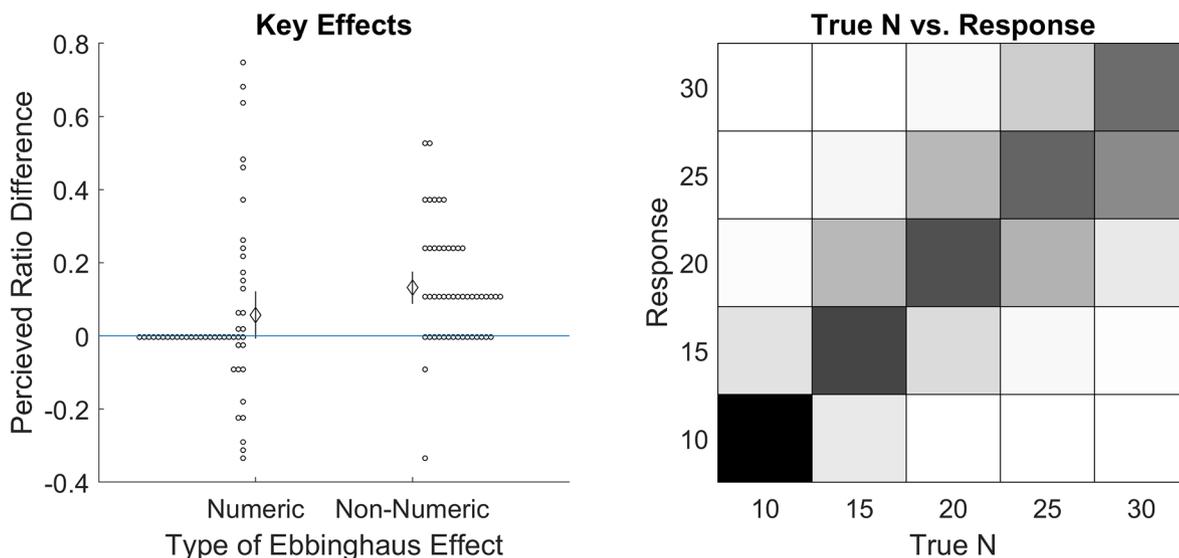
Note. The proportional numeric Ebbinghaus effect was again significantly above zero on average but not significantly different from the non-numeric effect (left). Responses again mainly fell along the diagonal containing correct responses (right).

General Discussion

Participants were asked to estimate the number of dots in arrays surrounded by context circles of different sizes. As in a previous paper (Picon et al., 2019), the participants gave higher numeric estimates when the arrays were surrounded by smaller context circles. In other words, the same classic stimulus manipulation that makes a single dot appear larger in the Ebbinghaus illusion also makes an array of dots appear more numerous. Controls in the analysis suggest this was not just a constant increase, which could instead be explained by the need to increase the number of context circles to keep a full surround. The use of an estimation method, rather than comparison, also guards against simple response biases crossing over from different forms of non-numeric perception (Picon et al., 2019). To statistically test the relation between the numeric and non-numeric Ebbinghaus-illusions, the present study also compared the numeric and non-numeric illusions in terms of their magnitude and found no significant difference. In other words, the same context circles create similar proportional increases in both perceived area and perceived number. This makes it possible that the two versions of the illusion reflect overlapping cognitive processes.

The interpretation, as agreed before the data were collected, has three aspects. First, it confirms a specific prediction from a density-area mechanism. These theorized mechanisms estimate number by scaling a density estimate with an area estimate (i.e. density times area equals number). In that case, whenever we increase perceived area, we should also increase perceived number. This prediction was positively confirmed here twice. Second, it suggests the mechanism may even use an unmoderated area estimate. The numeric and non-numeric effects were not merely in the same direction but also had no significant difference in their magnitudes. This could indicate that perceived area feeds into perceived number without any kind of filter or moderator between – or at least, any moderation is too small for present methods to detect. Third, the present results speak strongly against discrete theories of numeric perception for this kind of task. Such theories begin by stripping objects of their continuous properties, abstracting them as discrete entries, and then non-verbally counting the entries (Cordes et al., 2001; Meck & Church, 1983). This kind of theory could never predict illusory changes based on continuous aspects of surrounding contexts since all of that is stripped away in the pre-counting processing. The overall effect is to further build a case that this kind of numeric perception (adults, rapid, larger sets, visual-to-verbal estimation) relies on a density-area mechanism.

Taking a step back slightly, these findings could be important even outside the particular issue of a density-area mechanism. Even if density-area mechanisms were eventually rejected as sound theory, the existence of the numeric Ebbinghaus effect would still provide a simple bright-line criterion for future alternative theories. Any such theory would need to somehow respond to the changes in context circles with a proportional change in numeric perception. This must necessarily restrict the possibilities. However, from a different point of view, this is also the most pressing limitation of the present evidence: while the changing context circles here do lead predictably to changes in perceived area, they also change a variety of low-level visual statistics. While I am not aware of any theory that predicts the numeric Ebbinghaus effect from factors outside a density-area mechanism, it is entirely possible that one will become prominent in the future. In other words, the evidence here confirms a specific prediction from a density-area mechanism rather than demonstrating that theory unequivocally.

The comparison between the two experiments here does not particularly support existing theories around different mechanisms for sparse and dense arrays (Anobile et al., 2014; Cicchini et al., 2016), though this needs to be interpreted with caution. To start with, it is not necessarily clear that such a theory would always predict a difference in the specific illusion examined here. Beyond that, it is not necessarily clear how big of an effect this difference would be. Without some information along those lines it is difficult to evaluate whether the present study had adequate statistical power for those purposes. Please note that the present study was designed to test individual effects from previous studies with an estimated effect size of $d = 0.95$ rather than pursuing small differences across experiments.

The present findings, especially with sparser stimuli, are at odds with previous results (Cicchini et al., 2016) that suggest a disparity in sensitivity to number, area, and density in sparser arrays. As this is not the point of the paper here I will keep these comments limited. Of the three basic claims made by this previous work, the most obviously relevant for purposes here is the final one: when number sensitivity is predicted from density sensitivity and area sensitivity (assuming a density-area mechanism), the predicted thresholds for number are higher than the measured thresholds for number. This would seem to indicate that number judgements have a sensitivity that cannot be inherited from

density and area, thus ruling out a density-area mechanism. From my point of view, the obvious question is whether this particular pattern, this specific problem with the hierarchy of sensitivity to area/density/number, only appears in comparison tasks with conflicts between these three dimensions. If so, it might reflect something about differing ability to suppress response biases rather than differences in perceptual sensitivity – which could be very interesting but is not an argument against a density-area mechanism. It would be very helpful to further look into these issues as it is likely to provide both further insights into the density-area debate and more broadly into number perception.

It is also worth pointing out that the estimated effect sizes here (d = 0.45, 0.25) are (substantially) smaller than the effect sizes in previous papers (d = 0.95; Picon et al., 2019). There could be several reasons for this. First, this could be attributed to a shift in emphasis from scrutinizing the numerical Ebbinghaus effect to honing in on its proportional dimension. Second, it could be due to incidental differences in the screens and settings used by participants. Third, it could reflect the biases inherent in the publication system as it is actually very normal for registered replication efforts to return smaller effect sizes (Ioannidis, 2008). One way or another, it is likely appropriate for any future effort examining this illusion to adjust the sample size.

It may also be worthwhile to comment briefly on the unusual distribution of the observed numeric effects: many were very near zero in both experiments. To be more precise, the median effect was much closer to zero than the mean effect. There are at least three possible reasons for this. First, it might reflect some kind of individual differences. There may be some people who don't experience this illusion for some reason. Second, it may be an unexpected mathematical artifact. The calculation of the effect here involves dividing two sums with normally distributed error terms. Depending on the means and the standard deviations, this could result in a sampling distribution more like a Cauchy than a Gaussian. This would explain both the unusual shape (a Cauchy has a strong central 'peak' and 'fat tails') and the multiple extreme outliers that were trimmed. The third option, of course, is that it may have just been a fluke that would not replicate in future studies. Any set of real data will always have some unlikely spurious feature. Further research would be required to be certain.

As the final comment on the present data, the absent correlation between the numeric and non-numeric effects remains unexplained at this point. Obviously, it would have been convenient for the interpretation here if the two showed a robust correlation. Among explanations that do not deny a density-area mechanism, the simplest is that this just reflects inadequate precision to detect such a correlation. The observed correlation between two scores can only be as high as the square root of the product reliability – and even that can only be reached with a perfect correlation between the underlying constructs. The present study was designed primarily to detect differences in means. It remains unclear if an alternative design would detect a correlation between the numeric and non-numeric Ebbinghaus effects.

One of the more obvious tasks for future research is to break down this illusion and gain a more detailed understanding of the necessary and sufficient conditions for it. In the stimuli here, the smaller circles were more numerous (to keep a full ring) and closer. Things like the Delboeuf illusion (a simple ring that is nearer or further) could help clarify what aspects of the stimuli here help create the numeric effect. While the present study confirms that the numeric Ebbinghaus effect exists under present circumstances, it leaves open many questions about how to generalize this finding.

To conclude, the current study rigorously assessed the numerical Ebbinghaus effect, particularly its proportional version as predicted by a density-area mechanism, and found that it holds even under a more stringent testing regime. Across two experiments, participants estimated the arrays with smaller context circles to be approximately 6-8% more numerous on average than matching arrays with larger context circles. This roughly matches the size of the non-numeric Ebbinghaus effect with the same context circles (i.e. no significant difference in either experiment). This confirms a direct prediction from a density-area mechanism. It further gives no specific reason to posit a moderating process that limits the effect of changes in perceived area. This helps build the case that a density-area mechanism is the way that numeric perception of this type (adults, rapid, larger sets, visual-to-verbal estimation) occurs in humans. Further work will now be required to fully understand the circumstances that lead to this illusion and to square this observation with arguments against density-area mechanism theories.

Supplementary Materials

The Supplementary Materials include the following items (see Negen, 2023a, 2023b):

- A pre-registration which occurred after the preliminary data set (which was pilot work to see if the method was plausible) but before the two final data sets (which are the basis of the conclusions here).
- The full method, analysis code, and all research data

Index of Supplementary Materials

Negen, J. (2023a). *Area density mechanism for numeric estimation* [Pre-registration protocol]. OSF Registries. https://osf.io/e7mvh/

Negen, J. (2023b). *Area density mechanism for numeric estimation* [Full method, analysis code, and research data]. OSF. https://osf.io/4f9zs/

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science, 25*(1), 265–270. https://doi.org/10.1177/0956797613501520

Carey, S. (2009). *The origin of concepts.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications, 7*(1), Article 12536. https://doi.org/10.1038/ncomms12536

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. https://doi.org/10.3758/BF03196206

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences of the United States of America, 108*(49), 19552–19557. https://doi.org/10.1073/pnas.1113195108

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology, 18*(18), R855–R856. https://doi.org/10.1016/j.cub.2008.07.053

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668. https://doi.org/10.1038/nature07246

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

Jansen, C. J. M., & Pollmann, M. M. W. (2001). On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics, 8*(3), 187–201. https://doi.org/10.1076/jqul.8.3.187.4095

Leibovich, T., & Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *Quarterly Journal of Experimental Psychology, 67*(5), 899–917. https://doi.org/10.1080/17470218.2013.837940

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, Article E164. https://doi.org/10.1017/S0140525X16000960

MathWorks. (2021). *MATLAB* (No. 2021b) [Computer software]. https://uk.mathworks.com/products/matlab.html

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), 320–334. https://doi.org/10.1037/0097-7403.9.3.320

Paul, J. M., van Ackooij, M., ten Cate, T. C., & Harvey, B. M. (2022). Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nature Communications, 13*(1), Article 1340. https://doi.org/10.1038/s41467-022-29030-z

Picon, E., Dramkin, D., & Odic, D. (2019). Visual illusions help reveal the primitives of number perception. *Journal of Experimental Psychology: General, 148*(10), 1675–1687. https://doi.org/10.1037/xge0000553

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19*(6), 607–614. https://doi.org/10.1111/j.1467-9280.2008.02130.x

Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception, 34*(7), 847–856. https://doi.org/10.1068/p5273

Stoianov, I., & Zorzi, M. (2012). Emergence of a "visual number sense" in hierarchical generative models. *Nature Neuroscience, 15*(2), 194–196. https://doi.org/10.1038/nn.2996

Sullivan, J., & Barner, D. (2013). How are number words mapped to approximate magnitudes? *Quarterly Journal of Experimental Psychology, 66*(2), 389–402. https://doi.org/10.1080/17470218.2012.715655

Appendix: Summary Table

Question	Does the numeric Ebbinghaus illusion work in a way that supports a density-area mechanism for numeric perception? *Numeric Ebbinghaus Illusion:* When surrounded by smaller context circles, an array of dots appears more numerous. *Density-area mechanism:* Number is estimated by scaling a density estimate by an area estimate.
Hypothesis	The magnitude of the proportional numeric Ebbinghaus effect will be (a) above zero and (b) similar in magnitude to the non-numeric Ebbinghaus effect. *Proportional numeric Ebbinghaus effect:* a percentage increase in perceived number, as would arise directly from an increase in a scaling factor.
Sampling Plan	50 participants from Prolific with minimal restrictions: (a) normal or corrected-to-normal vision so they can see the stimuli clearly and (b) English fluency so that they can understand the task and ethics documents.
Analysis Plan	A formal model of the task will be applied to each participant to extract the estimated magnitude of the proportional numeric Ebbinghaus effect. These will be tested against zero with a one-tailed *t*-test and against the non-numeric effect with a two-tailed test. Exclusions will be based on overall performance (r<0.3) or large deviation from the mean (z>2.5).
Sensitivity Rationale	Previous research observed d = 0.95. 50 participants will provide over 99.9% power. Failure to show a difference against both benchmarks (zero, non-numeric effect) will lead to study withdrawal, guardrailing sensitivity.
Interpretations	If the proportional numeric Ebbinghaus effect is A. Not above zero and below non-numeric: Area perception has limited (perhaps zero) input into numeric perception of this type. B. Above zero but below non-numeric: consistent with a moderated density-area mechanism. C. Above zero and not different from non-numeric: consistent with a density-area mechanism, perhaps even an unmoderated version.
Theories that could be shown wrong	B or C outcome: speaks strongly against any theory of numeric perception that begins by stripping away context or continuous features. A outcome: speaks against theories of numeric perception that take perceived area as a major scaling input.