



Conjoint measurement of physical size and numerical magnitude: Numerals do not automatically activate their semantic meaning

Daniel Fitousi¹

Accepted: 25 July 2021 / Published online: 28 September 2021
© The Psychonomic Society, Inc. 2021

Abstract

I applied the methodology known as maximum likelihood conjoint measurement (MLCM) to diagnose how numerical magnitude affects the perception of physical size of numerals in the size congruity effect (SCE). Traditional theories of the SCE argue for automatic activation of numerical magnitude, and therefore the obligatory interaction in perception between number and size. Participants in the current study were presented with pairs of numerals varying on physical size and numerical magnitude. They judged which member of the pair was physically larger, while ignoring numerical magnitude. Three nested observer models (i.e. independent, additive, and saturated) of the contribution of physical size and numerical magnitude to perceived size were fit to the data using maximum likelihood. The independent observer model exhibited the best fit for the majority of observers. These results cast doubts on the view that numerical magnitude is activated automatically in judgments of physical size of numerals.

Keywords Size congruity effect · Numerical cognition · Conjoint measurement

Introduction

In the lofty realm of mathematics, numbers are abstract objects (Dantzig, 1954). In the real world, however, numbers come dressed in physical attributes. For example, numbers are written on road signs, appear on screens, or sounded through speakers in the airport. Thus, the perception of numbers in the empirical world involves both semantic and physical dimensions, and the semantic meaning cannot be processed without first perceiving the physical aspects. How do people perceive the *physical* dimension of a number? Does numerical magnitude affect perception of the numeral's physical size? The emerging consensus is that it does (see, Dehaene, 1997 for review), leading many researchers to argue that the semantic dimension (numerical magnitude) is processed automatically whenever a digit is presented for view for any purpose. An interaction between physical and numerical dimension of a number is therefore inevitable. The present study subjected the assump-

tion of automatic activation of numerical magnitude to a novel scrutiny, harnessing a rigorous psychophysical technique known as *maximum likelihood conjoint measurement* (MLCM, Ho, Landy, & Maloney, 2008, Knoblauch & Maloney, 2012). The technique is based on the classic conjoint measurement methodology originally developed by Luce and Tukey (1964), and its subsequent elaborations (Krantz, Luce, Suppes, & Tversky, 1971). It allows a rigorous examination of the degree to which numerical magnitude contributes to judgments of numerals' physical size. To anticipate the conclusions, I show that numerical magnitude does not interact with physical size. Consequently, I challenge the assumption of automatic perception of numerical magnitude.

The size congruity effect

When presented with pairs of numerals that vary on both physical size and numerical magnitude, people are faster and more accurate to select the *physically* larger member of the pair in congruent ($7-2$) than incongruent ($7-7$) condition. This phenomenon has been dubbed the *size congruity effect* (SCE). It entails that numerical magnitude intrudes on judgments of physical size, even when it is irrelevant to the task at hand. The SCE has been recorded in numerous studies and with different notations

✉ Daniel Fitousi
danielfi@ariel.ac.il

¹ Ariel University, Ariel, Israel

of number (Dehaene, 1997; Algom, Dekel, & Pansky, 1996; Besner & Coltheart, 1979; Schwarz & Heinze, 1998; Henik & Tzelgov, 1982; Tzelgov, Ganor-Stern, & Maymon-Schreiber, 2009; Tzelgov, Meyer, & Henik, 1992; Pansky & Algom, 1999; Fitousi, 2014; Fitousi, 2010; Fitousi, Shaki, & Algom, 2009; Fitousi & Algom, 2006; Fitousi & Algom, 2018). The pervasiveness of the SCE has led many researchers to interpret it as strong evidence for mandatory processing of numerical magnitude (Tzelgov et al., 1992). The current study challenges this view.

The evidence for automatic activation of number seems strong, but there are pervasive biases in the routine experimental design of the SCE that cast doubts on mandatory activation of numerical magnitude in the SCE. Algom et al. (1996) have identified two critical biases prevalent in SCE studies. First, there is an asymmetry in the number of stimuli used for the numerical and physical dimensions. Researchers usually use the numbers 1 to 9 for numerical magnitude, but only two or three values (small, medium, larger) for physical size. This asymmetry can determine the observed interaction (Algom et al., 1996; Pansky & Algom, 1999). Second, the relative discriminability of values along the number and size dimensions is not matched. Usually, the numerical dimension is more salient than the physical dimension. This results in interference from the more discriminable (number) to the less discriminable (size) dimension (Melara & Mounts, 1993; Pansky & Algom, 1999; Algom & Fitousi, 2016). When physical size was made more salient than numerical value, a reverse SCE emerged (Fitousi & Algom, 2006). Moreover, when number and size were matched on discriminability the SCE was vanished (Algom et al., 1996). The malleability of the SCE casts doubt on the automatic nature of numerical information processing.

Garner (1974) has distinguished between pairs of *integral* and *separable* dimensions. The former are processed together as a Gestalt and are difficult to tear apart, the latter can be readily separated in perception and maintain a certain degree of independence (Algom & Fitousi, 2016). The assumption of automatic activation also corresponds with the idea the number and size are integral dimensions. The studies of Algom and his associates (Algom et al., 1996; Pansky & Algom, 1999; Fitousi & Algom, 2006) have shown that numerical magnitude and physical size are separable dimensions, and that the SCE can result from known biases in the routine experimental design, not from truly integral processing. The present study further sought to test the level of independence of the number and size dimensions by harnessing a powerful non-speeded methodology—the MLCM (Ho, Landy, & Maloney, 2008; Qi, Chantler, Siebert, & Dong, 2015). The MLCM has been widely applied to classic psychophysical dimensions such color dimensions of hue,

chroma, and lightness (Gerardin, Dojat, Knoblauch, & Devinck, 2018; Rogers, Knoblauch, & Franklin, 2016), and surface gloss and bumpiness (Ho et al., 2008; Qi et al., 2015). The current study expands the scope of applications to physical and semantic dimensions of numerals.

The MLCM provides a stronger test of the automaticity approach than is afforded by the SCE. First, the MLCM is a theory-driven approach that is based on rigorous mathematical derivations (Krantz et al., 1971; Luce & Tukey, 1964), while the SCE is an operational measure, whose presence or absence in the data can be readily affected by a myriad of experimental factors (Algom et al., 1996; Pansky & Algom, 1999; Fitousi & Algom, 2006). Second, the MLCM is more explicit theoretically because it specifies the form of relations between dimensions, while the SCE is only an empirical observation that is amenable to circular explanations. Third, the MLCM is applied to a continuum of dimensional values (Ho et al., 2008; Qi et al., 2015), whereas the SCE is derived from two categories (congruent vs. incongruent). Finally, together with the Garner paradigm (Algom et al., 1996), the MLCM serves as an additional powerful source for converging operations on the separability of dimensions (see for, Rogers et al. 2016, 2018).

Method

Observers

A total of 24 observers participated in this study. They were recruited from Ariel University pool of participants (F=19, M=5, mean age = 23.3, $sd = 1.59$). Observers were compensated with a course credit. All observers had normal or corrected-to-normal vision. All observers were naive to the goals of the study. The number of observers was determined based on a power analysis according to the following logic. In MLCM there are three possible models that can fit to the data of an observer (independent, additive, and saturated, see Model section). Under the null hypothesis, any of these three strategies has equal chance of being adopted by the observer ($\pi = 0.333$). Under the alternative automatic hypothesis, observers combine numerical magnitude and physical size in an interactive fashion. Thus, the saturated model should be more prevalent than expected by chance ($\pi = 0.60$). I used the ‘pwr’ package (Champely, Ekstrom, Dalgaard, Gill, Weibelzahl, & Anandkumar, 2018) in the open source software R (Core Team, 2017) to compute the sample size needed to detect such an effect, with a power of 0.8 and significance level of $\alpha = 0.05$. A sample size of N=22 was found to meet these requirements.

Stimuli and apparatus

The stimuli were white numerals placed over a black background. The stimuli were created with the Microsoft Paint software. All numerical characters were in font type ‘Arial’. Five font sizes were used: 26, 36, 48, 58, and 72. Five numerical magnitudes were chosen 1, 3, 5, 7, and 9. The five font sizes and the five numerals were combined in a full factorial fashion to yield 25 possible numerals (5 levels of physical size \times 5 levels of numerical magnitude = 25). Figure 1 shows the stimuli on a Cartesian grid. Let i denote the rows of this grid and j denote its columns. The stimuli matrix represents a full factorial combination of i levels of physical size and j levels of numerical magnitude. For a given level of numerical magnitude j , physical size is increasing as we move up the i th rows. For a given physical size level i , numerical magnitude is increasing as we move to the right j th columns.

Procedure and design

On each trial observers viewed 1 of the 325 possible pairs of numerals from those illustrated in Fig. 1 (including self-comparisons). Viewing distance was approximately 56cm. The two numerals were presented simultaneously one next to the other on the center of the screen. The observer’s task was to judge which of the two numerals was physically larger while ignoring numerical magnitude. Each pair was presented 3 times, making a total of 975 trials¹. Each trial started with a central fixation point presented for 200 ms, then the pair of numerals appeared for 400 ms and was removed from the screen. The observer indicated whether the numeral on the left or the numeral on the right appeared to him or her as physically larger by pressing one of two corresponding keys on the keyboard (‘z’ or ‘m’). The observer’s response initiated the next trial. The task was not speeded. The actual observer’s response was recorded and served as the main dependent variable of interest.

Model

Additive conjoint measurement permits us *simultaneously* to measure and model the contributions of both physical size and numerical magnitude dimensions to perceived size. The results are estimates of two functions: *perceived size* and *perceived magnitude* that capture these contributions. The psychophysical task required for application of MLCM is a comparative judgment task (Luce & Tukey, 1964; Knoblauch & Maloney, 2012). On each trial, observers are presented with a pair of stimuli and indicate which member of the pair is higher (e.g., larger) on a relevant dimension,

¹This is the smallest number of repetitions needed to produce reliable parameter estimates (Ho et al., 2008)

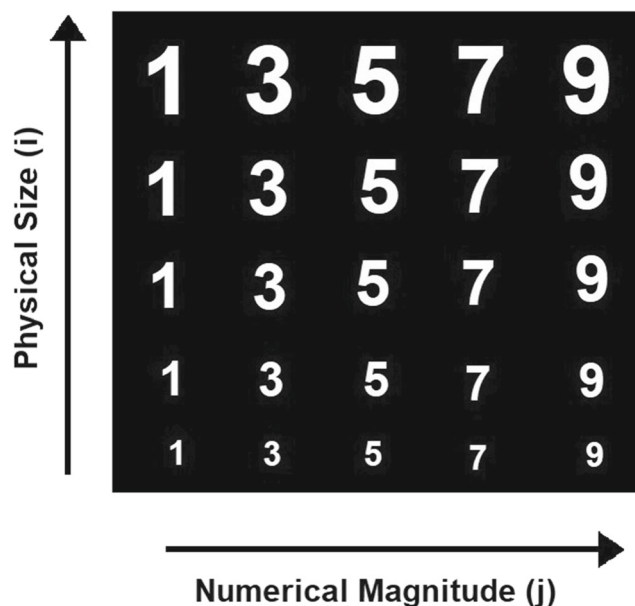


Fig. 1 The stimuli set used in the experiment. Five levels of numerical magnitude were combined with five levels of physical size. The i stand for levels of physical size and j for levels of numerical magnitude. Note that for a given level of numerical magnitude j , physical size is increasing up the i th rows. For a given level of physical size i , numerical magnitude is increasing as one moves to the right of the j th columns

while ignoring an irrelevant dimension. The procedure corresponds well with the comparative judgment task that has been typically used to gauge the SCE (Henik & Tzelgov, 1982). But one should bear in mind that it is not speeded, and that the comparisons also include pairs in which the physical sizes of the two members are comparable (sampled from the same row i of the Cartesian grid in Fig. 1).

In the current experiment, observers viewed all possible pairs of the 25 numerals (see Fig. 1), and on each trial judged which member of the pair was physically larger². Let ϕ^s denote values on the physical size scale and let ϕ^m denote values on the numerical magnitude scale. The value of physical size is constant across the i th row, and is denoted ϕ_i^s and the value of numerical magnitude is similarly constant in the j th column and is denoted ϕ_j^m . Comparison between two numerals can be represented as a comparison between two ordered pairs (ϕ_i^s, ϕ_j^m) and (ϕ_k^s, ϕ_l^m) . It is assumed that the visual system computes an estimate of perceived physical size that is based on both its physical size ϕ^s and numerical magnitude ϕ^m . The estimates of this quantity are $\psi^S(\phi_k^s, \phi_l^m)$ and $\psi^S(\phi_i^s, \phi_j^m)$, with uppercase letters for psychological measures and lowercase letters for actual physical or semantic variables. The resulting estimates of perceived physical size amounts to:

$$\Delta(i, j, k, l) = \psi^S(\phi_k^s, \phi_l^m) - \psi^S(\phi_i^s, \phi_j^m) + \varepsilon \quad (1)$$

²Self-comparisons were included, but they did not affect the model fitting. They are often incorporated to measure response bias

where the random variable $\varepsilon \sim \mathcal{N}(\mu, \sigma_\varepsilon^2)$ is a judgment error.

There are three nested models of the decision process that can be tested in the MLCM framework (Knoblauch & Maloney, 2012; Ho et al., 2008): (a) *the independent observer* (an ideal observer), (b) *the additive observer*, and (c) *the saturated observer*. Each model provides estimates of the perceptual scale values of physical size and how they are affected by values of numerical magnitude. Graphical illustrations of these models can be seen in Fig. 2. Because the models are well defined, they allow a rigorous test of the degree to which numerical magnitude intrudes on judgments of physical size. The independence (ideal observer) model assumes that decisions about physical size are made based only on physical size, with no contributions from numerical magnitude. The additive model assumes that decisions with respect to physical size are made based on the sum of component psychological responses elicited by physical size and numerical magnitude. The saturated model is based on the premise that decisions about physical size also include an interaction term that depends on the specific values of physical size and numerical magnitude, in addition to their simple additive combination.

If the SCE indeed reflects the automatic activation of magnitude due to strong interactive relations between number and size (Henik & Tzelgov, 1982; Tzelgov et al., 2009; Tzelgov et al., 1992), then the saturated model should provide the best fit to the data of all observers. In contrast, if the SCE results from pervasive biases in the standard experimental design, and the two dimensions are not truly integral, but rather separable dimensions (Fitousi & Algom, 2018; Algom et al., 1996), then the independent or additive models should provide the best fit to the data.

The three models form a nested series in which the independent model is the most constrained and the saturated model is the least constrained. A nested likelihood ratio test

is used to evaluate the models (Gerardin et al., 2018; Rogers et al., 2016; Ho et al., 2008; Fitousi, 2014)

The additive conjoint model replaces $\psi^S(\phi_i^s, \phi_j^m)$ by $\psi_i^{S:s} + \psi_j^{S:m}$ where $\psi_i^{S:s}$ is an additive contribution of physical size to psychological (or perceived) size that is constant across the i th row and $\psi_j^{S:m}$ is a comparable contribution of the numerical magnitude to perceived physical size that is constant across the j th column (see Fig. 1). $\psi_i^{S:s}$ and $\psi_j^{S:m}$ are parameters of the additive model that are estimated from data. The additive model is based on the assumption that numerical magnitude and physical size interact in determining perceived physical size, but that the contribution of a particular level of numerical magnitude $\psi_j^{S:m}$ to perceived physical size is independent of the physical size (Ho et al., 2008). The first term $\psi_i^{S:s}$ forms the psychophysical scale, mapping physical size to perceived size, and the second term $\psi_j^{S:m}$ represents the “contamination” of perceived physical size by changes in numerical magnitude. The assumption is tested here, along with the hypothesis that changes in numerical magnitude affect perceived physical size at all, which amounts to the hypothesis that $\psi_j^{S:m} = 0$ for all j .

The additive model can therefore be rewritten as:

$$\Delta(i, j, k, l) = \psi_k^{S:s} + \psi_l^{S:m} - \psi_i^{S:s} - \psi_j^{S:m} + \varepsilon \quad (2)$$

which can be rearranged to:

$$\Delta(i, j, k, l) = [\psi_k^{S:s} - \psi_i^{S:s}] + [\psi_l^{S:m} - \psi_j^{S:m}] + \varepsilon \quad (3)$$

The additive model for physical size is based on comparison of the perceived size levels of the two numerals with an additive contamination from the difference in perceived numerical magnitude. Because there are 5 levels along

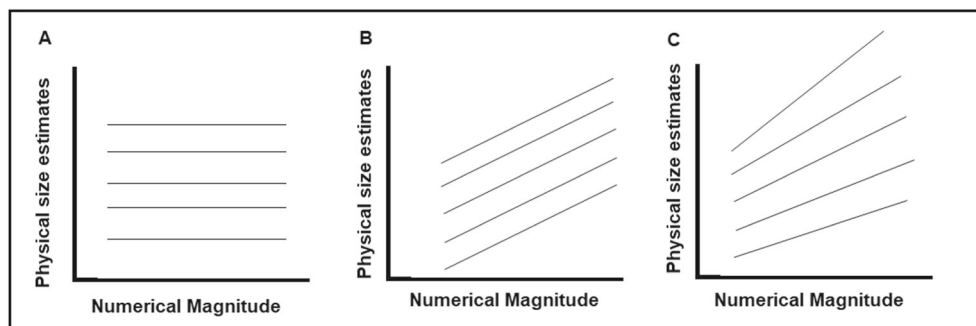


Fig. 2 Illustrations of the three hypothetical models of conjoint measurement with numerical magnitude and physical size: A. An independent observer, B. An additive observer, and C. A saturated observer. According to the independent observer the psychological response to physical size is affected only by changes in objective physical size with no contribution whatsoever from changes

in numerical magnitude. According to the additive observer psychological responses to physical size are also affected by additive contributions from numerical magnitude. Finally, according to the saturated model, psychological responses to physical size are determined by unique combinations of physical size and numerical magnitude values

each dimension, there are 2×5 levels plus 1 variance, which amount to 11 free parameters. To make the model identifiable, it is customary to fix the response at level 1 (the lowest) in each dimension to 0, $\psi_1^S = \psi_1^M = 0$, and the variance to 1. This reduces the number of free parameters to 8 (Knoblauch & Maloney, 2012; Ho et al., 2008).

The independent model is identical to the additive model except that there is no contamination of perceived physical size by numerical magnitude. The decision in this model is then based on:

$$\Delta(i, j, k, l) = [\psi_k^{S:s} - \psi_i^{S:m}] + \varepsilon \quad (4)$$

In this model the perceived difference in physical size depends only on the physical size of the numerals compared. Hence, the values of ψ_j^M are fixed at 0 and the total number of free parameters is 4.

The additive conjoint model assumes that the contamination of size by magnitude is additive and independent of size. Note that this assumption renders the additive model inconsistent with the automaticity approach (Tzelgov et al., 1992). The latter assumes that physical and numerical magnitudes are transformed onto a single abstract sensory code (Dehaene, 1997), and therefore predicts a stronger form of interaction than that predicted by the additive model. It is the saturated model that aligns best with the automaticity approach.

The saturated model includes an interaction factor that depends on the intensity levels of both physical size and numerical magnitude. The decision variable is defined according to:

$$\Delta(i, j, k, l) = [\psi_i^{S:s} + \psi_j^{S:m} + \psi_{ij}^{S:s:m}] - [\psi_k^{S:s} + \psi_l^{S:m} + \psi_{kl}^{S:s:m}] + \varepsilon \quad (5)$$

In this model, responses cannot be accounted for based on simple additive combination, but require the assumption of interactive terms. Hence, it is assumed that the response to each numeral is based on a unique combination of the separate contributions of physical size and numerical magnitude. Note that the 25 numerals that are presented in Fig. 1 are comprised of 5 levels of physical size and 5 levels of numerical magnitude. One cell in this grid is fixed at 0 leading to 24 free parameters. This maximal number of free parameters gives this model its name (saturated).

The three models comprise a nested set (Ho et al., 2008), where the independent model is the most constrained, the saturated model is the least constrained, and the additive model is at an intermediate level. The goals of the experiment are to estimate the perceptual scale values and model the contributions of both physical size and numerical magnitude from each observer's data, as well identifying the best fitting model.

Results

Accuracy

Data, codes, and other supporting materials can be downloaded from <https://data.mendeley.com/datasets/55x3jvs72f/1>. High level of accuracy is required for application of the MLCM (Ho et al., 2008), and this condition was met here. All observers performed at a level of 90% - 99% correct (mean = 96%). Note that in MLCM accuracy can be computed only for pairs of numerals with unequal physical size (76% of the trials). 24% of the pairs have the same value of physical size (i.e., sampled from the same row i), and therefore responses to those pairs cannot be designated as correct or incorrect. It should be noted that the input to the MLCM model is not accuracy rates, but rather the perceived ordering of pairs³.

To test for the presence of an SCE in the accuracy data, I divided the trials into congruent and incongruent trials. The SCE was computed as the difference in accuracy between congruent (0.97) and incongruent (0.96) trials, and it was not significant [$t(23) = 1.10, p = 0.13$] (one-tail). Its absence is not surprising, given that the design was completely balanced, with equal number of physical and numerical values (Algom et al., 1996).

MLCM

I used the MLCM package (Knoblauch & Maloney, 2014) in the open source software R (Core Team, 2017) to analyze the data. The default method uses the GLM function to estimate the model coefficients with a maximum likelihood criterion. The ML estimates are found with an iteratively reweighted least squares (IRLS) algorithm. The χ^2 statistic is computed from the differences of deviance of the nested model fits (see Knoblauch & Maloney, 2012 pp. 240-245).

I first compared the additive and independent models. I did it separately for each observer. The degrees of freedom for this specific test were computed as the difference between the coefficient estimates in each model (8 for the additive model - 4 for the independence model = 4). At a Bonferroni-corrected level of $p < .002$, the additive model provided better fit to the data than the independent model for only six out of 24 observers (obs. 2, 8, 9, 12, 19, and 21) (see Table 1). For two observers (obs. 6 and

³With enough trials the MLCM model fitting procedure can detect any deviation from the ideal (independent) observer model, including signatures of the saturated model. Knoblauch and Maloney (2012) explain that: "In any experiment, any pattern of response is consistent with any conjoint measurement representation or the absence of any representation. Even if the probability of choosing A over B is 0.999, there is a probability 0.001 of choosing B over A. Any pattern of response to any series of trials is possible, and we seek the explanation with greatest likelihood, not the only possible representation" (p.255)

Table 1 Comparison of independent and additive observer models for judgments of physical size

	Test Numerals (additive model x independent model)		
	DF	χ^2	p-value
Obs. 1	4	8.80	.07
Obs. 2	4	24.21	<.002***
Obs. 3	4	4.69	0.32
Obs. 4	4	2.65	0.61
Obs. 5	4	9.05	0.06
Obs. 6	4	14.37	0.006
Obs. 7	4	3.70	0.44
Obs. 8	4	75.67	<.002***
Obs. 9	4	43.16	<.002***
Obs. 10	4	0.73	0.94
Obs. 11	4	12.50	0.01
Obs. 12	4	13.40	<.002***
Obs. 13	4	8.58	0.07
Obs. 14	4	12.50	0.01
Obs. 15	4	5.65	0.22
Obs. 16	4	15.80	0.003
Obs. 17	4	5.62	0.23
Obs. 18	4	7.46	0.11
Obs. 19	4	47.01	<.002***
Obs. 20	4	9.68	0.04
Obs. 21	4	20.57	<.002***
Obs. 22	4	5.54	0.23
Obs. 23	4	2.69	0.61
Obs. 24	4	10.84	0.03

The Bonferroni-corrected level for multiple comparisons is $p < .002$.
***= $p < 0.002$

obs. 16) the additive model was on the verge of significance ($p=0.006$ and $p = 0.003$, respectively). This entails that for at least 75% of the observers, the independent (ideal) observer model provided a better description of the data. For these observers, judgments of physical size were immune to intrusions from numerical magnitude. It should be noted though that for 25% of the observers, the perception of physical size was contaminated by additive contributions from numerical magnitude. This implies that for those observers magnitude did affect size. I will refer to this result in the General Discussion.

I next derived the estimated contributions of physical size and numerical magnitude to perceived physical size. These estimates can inform us on how observers transformed the values of physical size and numerical magnitude onto a perceptual (psychological) scale of size. Figure 3 plots representative plots of perceptual scale values as a function

of the physical size values, separately for physical size and numerical magnitude. For physical size, observers exhibited the expected monotonic increase, entailing that observers indeed perceived numerals as physically larger as their physical size increased. The contribution of physical size to psychological size is granted, but does numerical magnitude contribute to judgments of psychological size? As can be noted in Fig. 3, additive contributions of numerical magnitude to psychological size are evident only in the data of eight observers (obs. 2, 6, 8, 9, 12, 16, 19, and 21). These are roughly the same observers whose data favored the additive over the independent model. For the remaining (majority) of observers, the contribution of numerical magnitude to perceived size was null. This is also in full agreement with the results of the model selection procedure. Interestingly, the magnitude scales varied across these observers, increasing in some, but decreasing in others. This suggests that increasing semantic magnitude does not in all cases increase the perception of physical size.

Next, I compared the saturated and the additive models (see Table 2). The χ^2 statistics in this case has 16 degrees of freedom because it is computed as the difference in number of free parameters of the less constrained model (the saturated model =24) and the more constrained model (the additive model=8). At a Bonferroni-corrected level of $p < .0083$, the saturated model exhibited better fit for only one of the observers (see Table 2). I also fit the saturated model to the data, separately for each observer's data. Figure 4 projects the estimated perceived physical size values against the numerical magnitudes according to this model. Each point corresponds to one of the 25 numerals in the experiment. The visual patterns reveal that the horizontal lines connecting the values at each magnitude level are parallel to each other for most observers. Eyeballing Fig. 4 supports the conclusions drawn from the modeling. For the participants whose data favored the additive observer model, the lines are parallel but increasing (obs. 2, 8, 12, 19) or parallel but decreasing (obs. 21).

In sum, out of 24 observers, 18 perceived physical size and numerical magnitude according to an independent (ideal) observer model, 5 observers according to an additive observer model and only one according to a saturated observer model. Even under a more liberal statistical approach, in which the additive model is as prevalent as the ideal observer model, the saturated model remains a rarity. The results are generally inconsistent with the predictions of the automaticity approach to number perception (Henik & Tzelgov, 1982; Tzelgov et al., 2009; Tzelgov et al., 1992). They are in line with the attention approach (Algom et al., 1996; Fitousi & Algom, 2018), whose proponents ascribe the SCE to pervasive biases in the routine SCE design, and not to truly abiding integrality of dimensions.

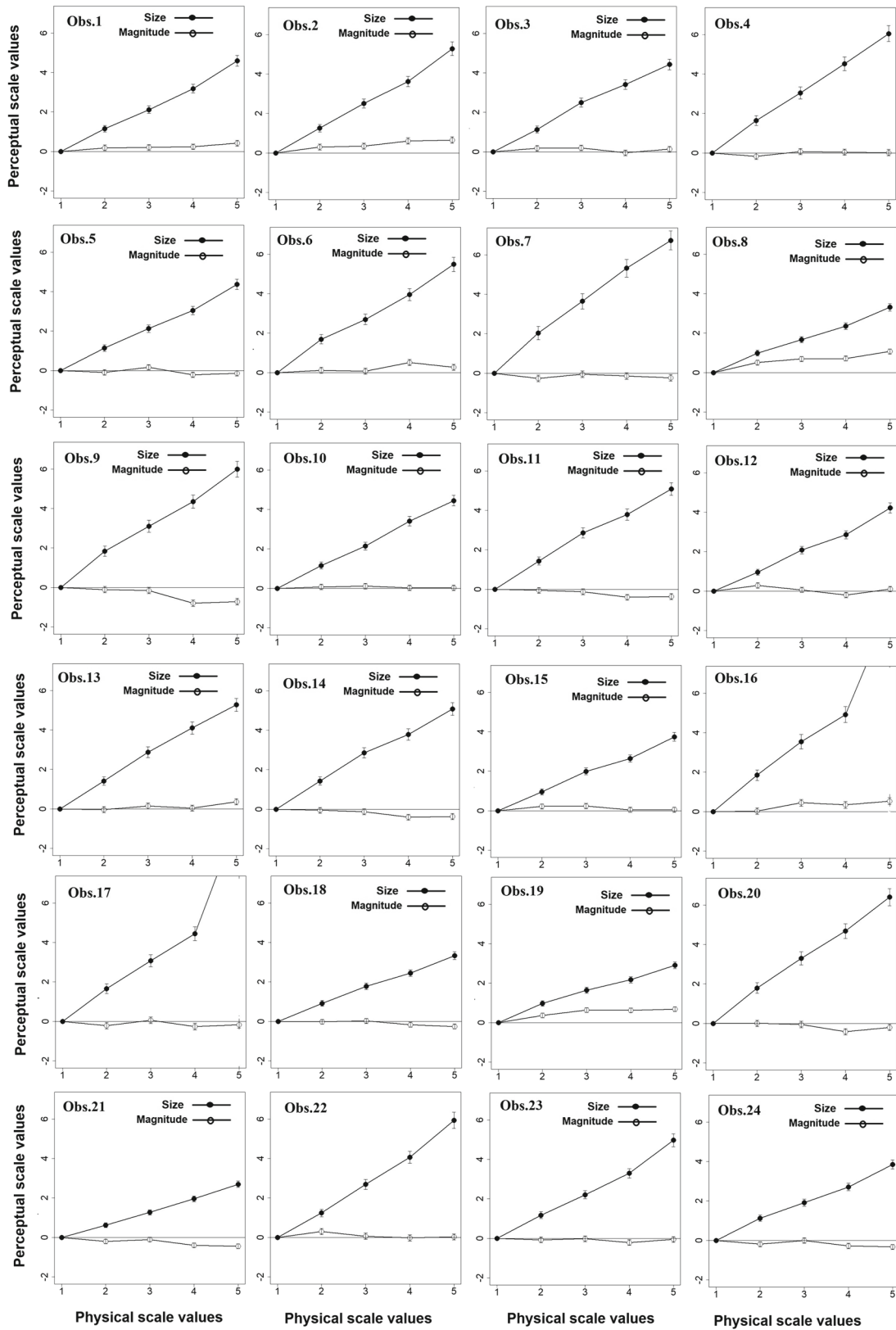


Fig. 3 Estimated scales for judgments based on physical size. Additive model average estimates for the 24 observers. The black circles connected by a line show the contribution of physical size and the white circles connected by a line show the contribution of

numerical magnitude. Error bars are standard errors of the mean, and are computed based on the Hessian of the maximum likelihood matrix (Knoblauch & Maloney, 2012)

Table 2 Comparison of additive and saturated observer models for judgments of physical size

	Test Numerals (saturated model x additive model)		
	DF	χ^2	p-value
Obs. 1	16	17.58	0.34
Obs. 2	16	26.77	0.04
Obs. 3	16	4.69	0.32
Obs. 4	16	24.84	0.07
Obs. 5	16	15.34	0.49
Obs. 6	16	23.45	0.10
Obs. 7	16	20.06	0.21
Obs. 8	16	13.76	0.61
Obs. 9	16	12.34	0.72
Obs. 10	16	16.33	0.42
Obs. 11	16	14.12	0.58
Obs. 12	16	22.99	0.11
Obs. 13	16	29.18	0.02
Obs. 14	16	14.12	0.58
Obs. 15	16	20.23	0.20
Obs. 16	16	35.32	0.003***
Obs. 17	16	20.37	0.20
Obs. 18	16	21.82	0.14
Obs. 19	16	20.37	0.20
Obs. 20	16	12.09	0.73
Obs. 21	16	6.99	0.97
Obs. 22	16	22.99	0.11
Obs. 23	16	12.63	0.69
Obs. 24	16	15.77	0.46

The Bonferroni-corrected level for multiple comparisons is $p < .0083$
 ***= $p < 0.0083$

Discussion

The SCE plays a central role in numerical cognition. It has been often interpreted as evidence for automatic activation of semantic magnitude and for the notion that magnitude and size are converted onto a common abstract code (Dehaene, 1997). It is therefore important to understand how the dimensions of physical size and numerical magnitude interact, and specifically whether and how the semantic dimension of magnitude contaminates the perception of size. The present study examined these questions directly, by applying the maximum likelihood conjoint measurement technique (Ho et al., 2008; Knoblauch & Maloney, 2012) to the dimensions of physical size and numerical magnitude. If the traditional automatic account were correct, then the MLCM should have favored the saturated observer model over the additive or (ideal) independent observer model. That was not the case. The present study showed that for most observers magnitude was perceived indepen-

dently of size, or affected size according to an additive observer model. The data of only one participant favored the saturated observer model. These results argue against an automatic activation of semantic magnitude, and also against the idea that number and size are transformed onto a common abstract code. The results are consistent with the attention approach advanced by Algom, Fitousi and their colleagues (Algom et al., 1996; Fitousi & Algom, 2018).

Independence of dimensions has already been shown by Algom and colleagues in the Garner paradigm (Pansky & Algom, 1999, 2002; Fitousi & Algom, 2006, 2020), where the dimensions have been found to be separable. Algom and colleagues have shown that the great majority of designs that measured the SCE, induce built-in experimental biases. For example, by picking two values for physical size and nine values for numerical values, experimenters have unwittingly induced an informational bias. Consequently, observers were attracted to the more variable dimension (magnitude), which resulted in SCE. Algom and colleagues have shown that the SCE can be generated, reversed, or eliminated at will through minor manipulations of such context (Algom et al., 1996; Pansky & Algom, 1999, 2002; Fitousi & Algom, 2006). For example, by using equal number of physical size and numerical magnitude values the SCE was eliminated. The outcomes from the conjoint measurement and Garner paradigms provide a set of converging operations on the separability of number and size (Algom & Fitousi, 2016; Garner, Hake, & Eriksen, 1956; Von Der Heide, Wenger, Bittner, & Fitousi, 2018). In addition, a recent study by Fitousi and Algom (2018) has applied the *System Factorial Technology* (SFT) (Townsend & Nozawa, 1995) – a powerful stochastic modeling methodology – to uncover the attendant architecture of processing magnitude and size. The results revealed the absence of an interactive architecture. Magnitude and size were processed according to a serial architecture and with a self-terminating stopping rule (meaning that magnitude was not obligatory for processing to halt).

Generalization of the results to other designs

The design employed in the current study is unique because: (a) it models comparative judgments of non-speeded responses rather than speeded responses, and (b) it is balanced in the sense that the stimuli set is comprised of an equal number of physical and numerical values. Thus, one may argue that the results obtained in this design may not generalize to other more ‘traditional’ designs, in which the SCE emerges (Henik & Tzelgov, 1982). In other words, one may claim that the dimensions of number and size should appear as dependent (or processed according to a saturated observer model) in traditional experiments. This argument is utterly misguided. First, separability (or

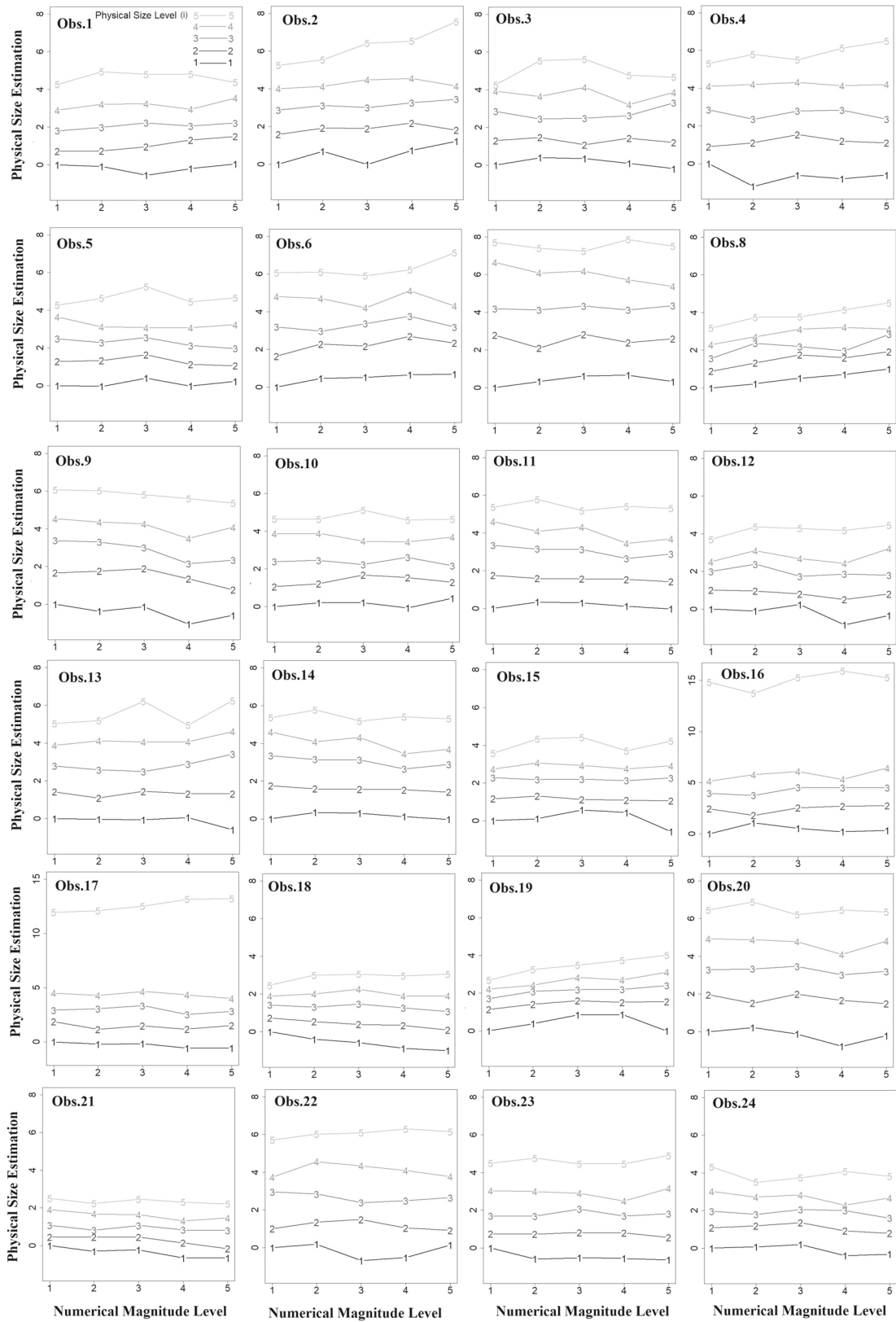


Fig. 4 Results of the saturated model with the estimated contributions for each combination of physical size and numerical magnitude for each observer. Levels of physical size are coded according to: (a)

numbers 1-5, with 1 being the smaller and 5 larger physical size, and (b) lines' color, which gradually shift from black (small physical size) to gray (large physical size)

integrality) is an intrinsic aspect of stimuli (Algom et al., 1996; Algom & Fitousi, 2016; Garner, 1974) and holds under different experimental contexts. Once the dimensions have been found to be separable or independent, their interaction in unbalanced designs can be safely ascribed to other experimental factors, not the stimuli themselves (see, Garner & Felfoldy, 1970). The upshot is that the emergence of an SCE does not imply that the dimensions of size and magnitude become dependent or integral in such designs. This point has been put succinctly by Algom et al. (1996): “Interactions, when they appear, are not intrinsic to dimensions involving numbers”. (p. 569). Pansky and Algom (1999) further claimed that: the additional sources of variations reside in the entire stimulus ensemble, not within the individual stimulus” (p.56).

Second, the independence of size and number cannot be decided based only on the presence or absence of SCE. The SCE is an empirical result not a theoretical account. In the absence of converging operation, any theoretical concept (e.g., numerical activation) is only a restatement of the empirical result (Algom & Fitousi, 2016). To avoid circular reasoning, Garner has noted the need for converging operations (Garner et al., 1956). Therefore, the separability of size and number is now supported not by one, but several converging operations from the: MLM, Garner (Algom et al., 1996; Pansky & Algom, 1999), and SFT (Fitousi & Algom, 2018) paradigms. All of which are well-established, rigorous, meta-theoretical tools.

Individual differences in the perception of size and number

A note is in order regarding the presence of individual differences in processing of magnitude and size. For 25% of the observers, the perception of physical size was contaminated by additive contributions from numerical magnitude. This implies that some participants were affected by numerical magnitude. One may argue that this result lends some support to the automatic approach. But as I have expounded at the outset, a model of additive ‘contamination’ is not predicted by the automatic approach (Dehaene, 1997). In particular, the accepted marker of automatic activation – the SCE – implies an interactive pattern by which an increase in numerical magnitude leads to a disproportional increase in the estimations of large-size numerals compared to small-size numerals. Moreover, for the portion of observers whose data favored the additive model, the effect of numerical magnitude on size was characterized by two opposite patterns. For some observers, increasing numerical magnitude led to a decrease in perceived physical size, while for others, the reverse was true. Most importantly, the data strongly rejected the saturated model, suggesting that number and size are

perceived as two independent entities, and that these entities are not transformed onto a common abstract code. In any event, the findings of individual differences in the way people handle numerical information opens the door for other future studies in numerical cognition on the presence of individual strategies.

Data Availability Data can be downloaded from <https://data.mendeley.com/datasets/55x3jvs72f/2>.

Open Practices Statement All data, codes, and supporting materials can be downloaded from here <https://data.mendeley.com/datasets/55x3jvs72f/2>.

References

- Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition*, 24(5), 557–572.
- Algom, D., & Fitousi, D. (2016). Half a century of research on Garner interference and the separability–integrality distinction. *Psychological Bulletin*, 142(12), 1352–1383.
- Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, 17(5), 467–472.
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., & Anandkumar, A. (2018). Package ‘pwr’. R package version, 1(2).
- Core Team, R. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: <https://www.R-project.org/>. [Google Scholar].
- Dantzig, T. (1954). Number, the language of science; a critical survey written for the cultured non-mathematician. Plume.
- Dehaene, S. (1997). The number sense: How the mind creates mathematics. OUP USA.
- Fitousi, D. (2010). Dissociating between cardinal and ordinal and between the value and size magnitudes of coins. *Psychonomic Bulletin & Review*, 17(6), 889–894.
- Fitousi, D. (2014). On the internal representation of numerical magnitude and physical size. *Experimental Psychology*, 61, 149–163.
- Fitousi, D., & Algom, D. (2006). Size congruity effects with two-digit numbers: Expanding the number line? *Memory & Cognition*, 34(2), 445–457.
- Fitousi, D., & Algom, D. (2018). A system factorial technology analysis of the size congruity effect: Implications for numerical cognition and stochastic modeling. *Journal of Mathematical Psychology*, 84, 57–73.
- Fitousi, D., & Algom, D. (2020). A model for two-digit number processing based on a joint Garner and system factorial technology analysis. *Journal of Experimental Psychology: General*, 149(4), 676–700.
- Fitousi, D., Shaki, S., & Algom, D. (2009). The role of parity, physical size, and magnitude in numerical cognition: The SNARC effect revisited. *Perception & Psychophysics*, 71(1), 143–155.
- Garner, W. R. (1974). The processing of information and structure. Psychology Press.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1(3), 225–241.
- Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, 63(3), 149–159.

- Gerardin, P., Dojat, M., Knoblauch, K., & Devinck, F. (2018). Effects of background and contour luminance on the hue and brightness of the watercolor effect. *Vision Research*, *144*, 9–19.
- Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395.
- Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*(2), 196–204.
- Knoblauch, K., & Maloney, L. (2014). MLCM: Maximum likelihood conjoint measurement. R package version 0.4.1.
- Knoblauch, K., & Maloney, L. T. (2012). Modeling psychophysical data in R (32). Springer Science & Business Media.
- Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). Foundations of measurement, vol. i: Additive and polynomial representations. New York Academic Press.
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*(1), 1–27.
- Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, *21*(5), 627–645.
- Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 39–58.
- Pansky, A., & Algom, D. (2002). Comparative judgment of numerosity and numerical magnitude: attention preempts automaticity. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28*, 259–274.
- Qi, L., Chantler, M. J., Siebert, J. P., & Dong, J. (2015). The joint effect of mesoscale and microscale roughness on perceived gloss. *Vision Research*, *115*, 209–217.
- Rogers, M., Franklin, A., & Knoblauch, K. (2018). A novel method to investigate how dimensions interact to inform perceptual salience in infancy. *Infancy*, *23*(6), 833–856.
- Rogers, M., Knoblauch, K., & Franklin, A. (2016). Maximum likelihood conjoint measurement of lightness and chroma. *JOSA A*, *33*(3), A184–A193.
- Schwarz, W., & Heinze, H. j. (1998). On the interaction of numerical and size information in digit comparison: a behavioral and event-related potential study. *Neuropsychologia*, *36*(11), 1167–1179.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*(4), 321–359.
- Tzelgov, J., Ganor-Stern, D., & Maymon-Schreiber, K. (2009). The representation of negative numbers: Exploring the effects of mode of processing and notation. *Quarterly Journal of Experimental Psychology*, *62*(3), 605–624.
- Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 166–179.
- Von Der Heide, R. J., Wenger, M. J., Bittner, J. L., & Fitousi, D. (2018). Converging operations and the role of perceptual and decisional influences on the perception of faces: Neural and behavioral evidence. *Brain and Cognition*, *122*, 59–75.

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