nature human behaviour

LETTERS

PUBLISHED: 17 JULY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0145

Selective overweighting of larger magnitudes during noisy numerical comparison

Bernhard Spitzer[1,2*], Leonhard Waschke[3] and Christopher Summerfield[1]

Humans are often required to compare average magnitudes in numerical data; for example, when comparing product prices on two rival consumer websites. However, the neural and computational mechanisms by which numbers are weighted, integrated and compared during categorical decisions are largely unknown[1–5]. Here, we show a systematic deviation from 'optimality' in both visual and auditory tasks requiring averaging of symbolic numbers. Participants comparing numbers drawn from two categories selectively overweighted larger numbers when making a decision, and larger numbers evoked disproportionately stronger decision-related neural signals over the parietal cortex. A representational similarity analysis[6] showed that neural (dis)similarity in patterns of electroencephalogram activity reflected numerical distance, but that encoding of number in neural data was systematically distorted in a way predicted by the behavioural weighting profiles, with greater neural distance between adjacent larger numbers. Finally, using a simple computational model, we show that although it is suboptimal for a lossless observer, this selective overweighting policy paradoxically maximizes expected accuracy by making decisions more robust to noise arising during approximate numerical integration[2]. In other words, although selective overweighting discards decision information, it can be beneficial for limited-capacity agents engaging in rapid numerical averaging.

Healthy humans ($n = 24$) viewed sequentially occurring symbolic numbers (samples: $n = 10$; range: 1–6, uniformly sampled) drawn from two categories, and were asked to indicate with a key press which category had the larger average (Fig. 1a). Categories were distinguished by their font colour (red versus green; visual condition) or the voice in which they were spoken (male versus female; auditory condition). Fully informative performance feedback followed each response. Discrimination performance (visual: $68.3 \pm 0.9\%$; auditory: $69.8 \pm 1.2\%$) did not differ between auditory and visual conditions (Wilcoxon signed-rank test: $P = 0.17$).

We used a simple psychophysical model to understand the rational policy for performing noisy numerical averaging in our task (see Methods). Model input X_i was the number occurring on each sample, i, normalized (for convenience) within the range −1 to 1. The model was parameterized to allow two potential sources of loss during averaging. The first, kappa (k), encoded a potential compression of the number line, allowing numbers to carry different weights in the decision: each sample X_i was transformed to a momentary decision value via a sign-preserving exponential function of the form $(X+b)^k$, where b is an additive offset parameter. When $k<1$, the transfer function has a sigmoidal form that compresses outlying values ($X_i \gg 0$ or $X_i \ll 0$) relative to inliers ($X_i \approx 0$; Fig. 1b, light grey lines). The converse is true when $k > 1$ (see Fig. 1b, dark grey lines). The second source of loss was assumed to occur after the processing of each sample; that is, during numerical averaging or at the response itself[7,8]. To generate simulated model choices, we passed the difference in cumulative decision values for each category through a sigmoidal function with inverse slope sigma (s), where higher values of s (that is, low slopes) indicate more noise in neural computation.

We then used our simulations to explore how the accuracy-maximizing policy changes under different values of compression, k, and decision noise, s. In the absence of noise (for example, in perfect averaging), the optimal policy is to leave the numbers uncompressed ($k = 1$); other policies discard numerical information before averaging (Fig. 1c, dark grey lines). However, as integration noise increases ($s \gg 0$), the accuracy-maximizing value of k increases (Fig. 1c, light grey lines); in other words, accuracy is maximized by giving more weight to outlying numbers (for example, 1 and 6) than inlying numbers (for example, 3 and 4), just as 'selective integration' has been shown to maximize accuracy in the presence of higher integration noise[2]. This was the case both under no bias ($b = 0$; Fig. 1c) and under a bias towards overweighting larger numbers ($b > 0$; Fig. 2). In the latter case, the accuracy-maximizing policy gives especially high weight to large outlying numbers (for example, 6; right panel in Fig. 1b). In both cases, during noisy numerical averaging (for example, when capacity is limited and integration is leaky or imperfect), the best policy is to base choices principally on more extreme (outlying) values in the numerical sequence (that is, $k > 1$).

Turning to the human data, we examined choice probabilities to estimate the influence of each sample (numbers 1–6) on the decision (Fig. 1d). In terms of absolute decision weight, in both the auditory and visual tasks, participants overweighted the higher numbers, 5 (relative to 2; Wilcoxon signed-rank tests: both $P < 0.001$) and 6 (relative to 1; both $P < 0.001$), when making their choices. Furthermore, the weight functions deviated significantly from linearity (visual: $F_{5,115} = 16.60$; auditory: $F_{5,115} = 17.35$; both $P < 0.001$). This behaviour was captured by fitting the psychophysical model to the human data, maximizing the likelihood of choices at the single-trial level (purple dots). Estimated values of s (integration noise) averaged about 1.8 (see below), a value at which values of $k > 1$ maximize accuracy (Fig. 1c; see Fig. 2a for detailed simulations). Consistent with this policy, the obtained best-fitting values of k (Supplementary Fig. 1a) significantly exceeded 1 (visual: $k = 1.95 \pm 0.19$; auditory: $k = 1.91 \pm 0.18$; Wilcoxon signed-rank tests: both $P < 0.001$; difference between conditions: $P = 0.80$), indicating an overweighting of outliers with a positive offset bias (visual: $b = 0.44 \pm 0.07$, $P < 0.001$; auditory: $b = 0.27 \pm 0.06$,

[1]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK. [2]Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany. [3]Department of Psychology, University of Lübeck, 23562 Lübeck, Germany.
*e-mail: bernardodispitz@gmail.com

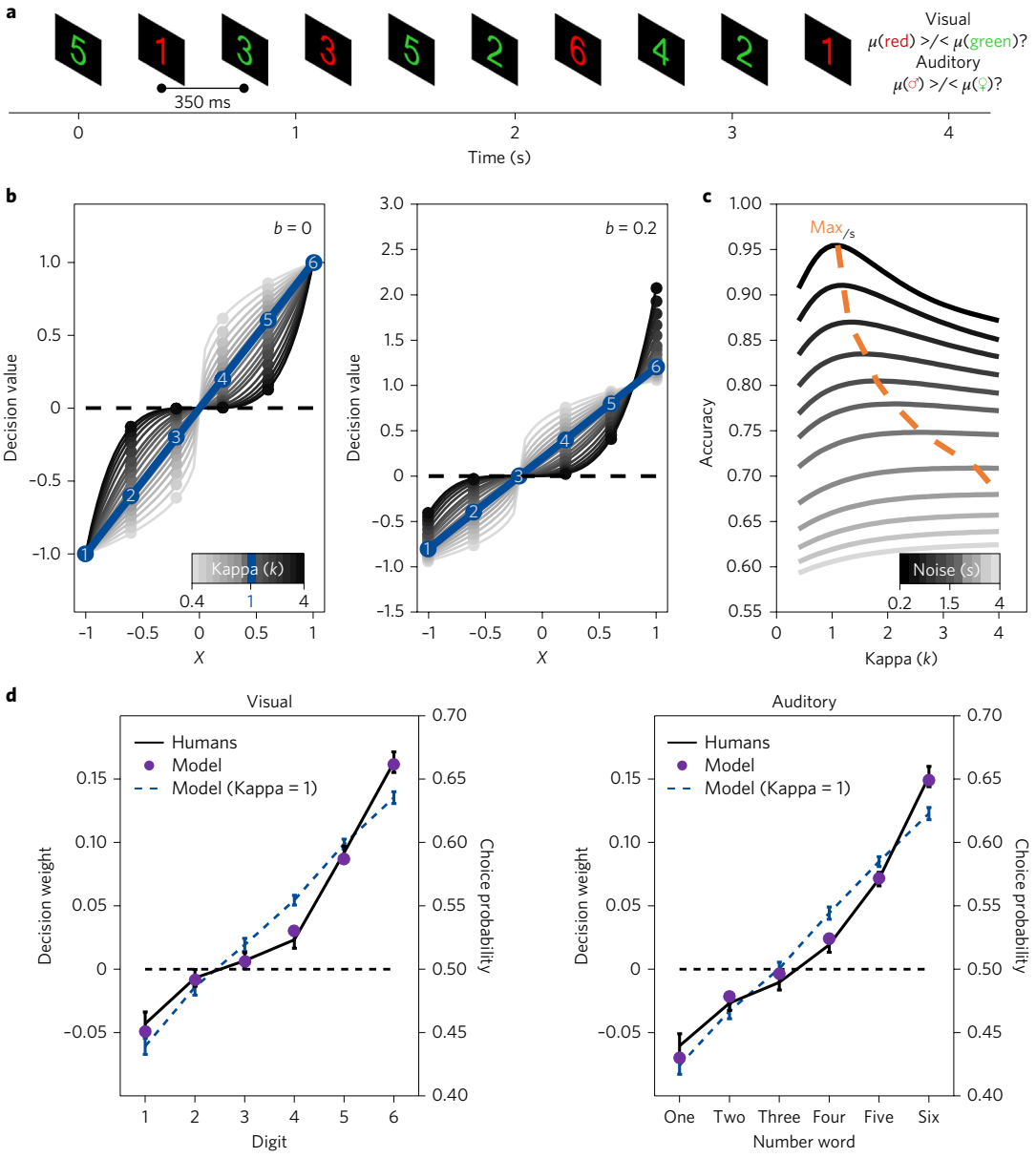

Figure 1 | Task, model simulations and human behaviour. a, Example trial sequence from the visual task. Ten numbers appeared in red or green font separated by 350 ms. The task was to report whether the average (μ) of the red or green numbers was higher (see right). In the auditory task, participants compared the average of numbers spoken in a male or female voice. **b**, Function mapping of sensory inputs, X, onto a decision value, $DV = (X + b)^k$ for different values of k (light grey lines, $k < 1$; dark grey lines, $k > 1$) and $b = 0$ (left panel) or $b = 0.2$ (right panel). **c**, Predicted accuracy under different values of k (x axis) and integration noise, s (lines), where the light grey lines correspond to larger values of s (that is, noisier decisions; greyscale: s is increased monotonically from 0.2 to 4). Simulations are shown for $b = 0$ (see Fig. 2 for simulations with $b > 0$). The orange line indicates the values of k that lead to maximum accuracy given noise level s ($\max_{/s}$). **d**, Left panel: decision weights for numbers 1–6 in the visual task. The black line shows human data ($n = 24$) and the purple dots show the predictions of the best-fitting model with $k = 1.95$ and $s = 1.75$ (mean estimates over subjects). The blue dashed line shows the fitted model predictions for $k = 1$. The decision weight (left y axis) is expressed as 'choice probability – 0.5'. Choice probabilities (right y axis) are inferred from the relative frequency of the samples' category (red/green) being chosen at the end of the trial. The error bars show s.e.m. Right panel: same as the left, but for the auditory condition.

$P < 0.001$), which was slightly greater in the visual condition ($P = 0.013$), confirming the preference for large numbers (for example, 6) over small ones (for example, 1). Note that 'anti-compression' for large numbers is the opposite of what would be expected from scalar variability; that is, if numbers were weighted according to Weber's law[9–12].

For comparison, the weights from an equivalent simulated observer with $k = 1$ (no compression) are shown in blue (Fig. 1d, dashed lines). A quantitative model comparison indicated that this model fit the data more poorly (Wilcoxon signed-rank test on Akaike information criterion (AIC) values: visual, $P = 0.002$; auditory, $P = 0.004$). Quantitative analysis also ruled out a model in which participants simply 'counted' the larger numbers (see Supplementary Information and Supplementary Fig. 1c). The introduction of one further parameter encoding a leak in the integration process allowed the psychophysical model to capture the full pattern of decision weighting as a function of sample position (1–10) and numerical value (1–6) (that is, 60 data points with 5 parameters;

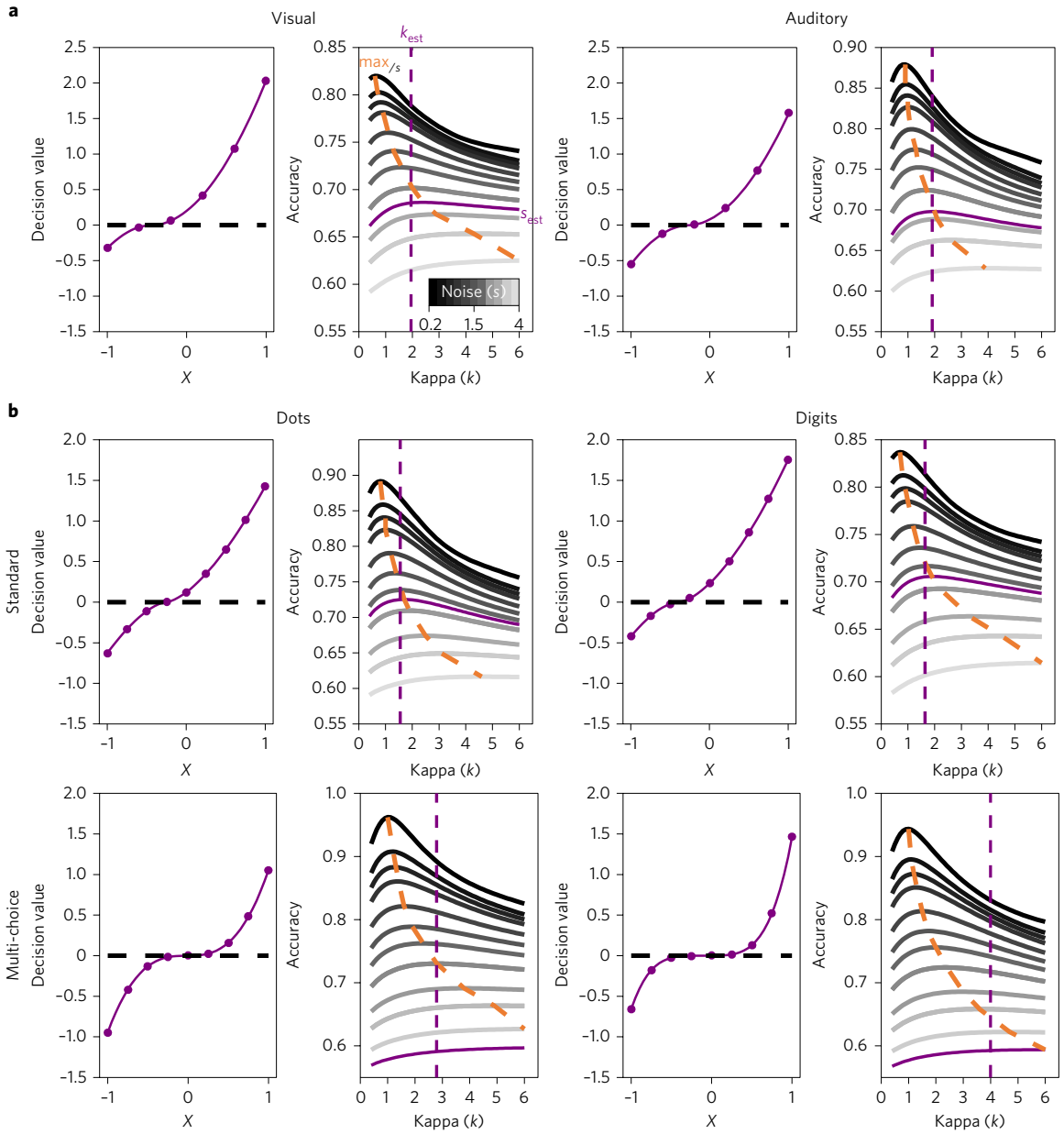

Figure 2 | Overview of model results and simulations for each experiment and condition. Left panels illustrate best-fitting mapping function (*x* axis: sensory input, *X*; *y* axis: decision value, DV). Right panels show simulated model accuracy (same conventions as in Fig. 1c) under the best-fitting parametrization in humans. Human kappa (k_{est}, dashed vertical) and noise level (s_{est}, solid) for each condition are highlighted in purple. **a**, Main experiment (see Fig. 1a,d; $n = 24$). **b**, Supporting experiment (see Supplementary Fig. 2; $n = 21$). In all conditions, k_{est} was increased (>1) towards the maximum predicted accuracy under the estimated noise level (s_{est}). Note that in cases where k_{est} fell short of the theoretical maximum (dashed purple versus dashed orange), the associated differences in predicted accuracy were relatively minor (<1%).

Supplementary Fig. 1b). Inclusion of the leak both reliably improved the overall fits (Wilcoxon signed-rank tests on AIC values: both $P < 0.001$) and reduced the best-fitting estimates of s (visual: 1.11 ± 0.08 versus 1.75 ± 0.12; auditory: 1.29 ± 0.09 versus 1.85 ± 0.12; Wilcoxon signed rank tests: both $P < 0.001$), suggesting that imperfect memory is itself a contributor to the cost of integration. Statistical tests with model and human as fixed factors showed no overall differences or interactions with experimental factors (all $F < 3$, all $P > 0.05$, corrected), confirming the ability of the model to capture human performance.

These findings suggest that during numerical averaging, decision values are 'anti-compressed' in precisely a way that will compensate for 'late' noise in the integration process and consequently maximise rewards (Fig. 2a). To directly test whether human decision policies adapt to the level of late noise in the task, we conducted a new experiment in which the cost of integration was manipulated directly in two distinct conditions. A fresh cohort of participants ($n = 21$) viewed sequential number samples (Supplementary Fig. 2a; red and green digit or dot displays, $n = 8$ samples, range 1–9) with instructions to compare averages along a single axis (for example, red versus green) or multiple axes (for example, both red versus green and digits versus dots; see Supplementary Information for details). Fitting these data with the psychophysical model described above indicated that 'late' noise was indeed lower in the single-axis condition (the 'standard' task: $s = 1.33 \pm 0.22$) than in the multiple-axis condition ('multi-choice' task: $s = 4.80 \pm 0.82$)

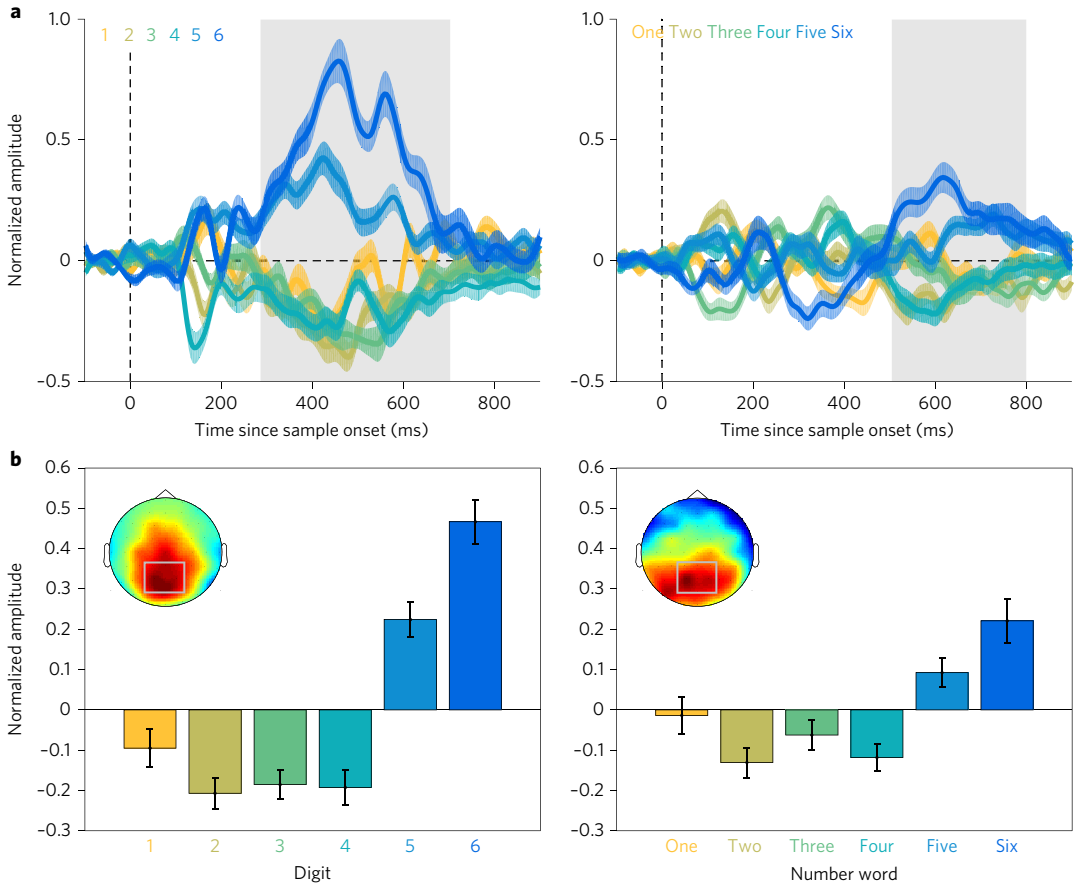

Figure 3 | CPP analysis. Left panels: visual; right panels: auditory. **a**, Mean-subtracted EEG signals ($n = 24$) evoked by numbers 1–6 over centro-parietal electrodes, averaged across sample positions. The coloured shaded regions show s.e.m. The grey shaded area indicates the time window of significant CPP modulations identified by leave-one-out permutation (see Methods). **b**, Mean CPP amplitudes in the time windows identified in **a**. The error bars show s.e.m. The insets illustrate scalp topography for sample number 6.

with a significant difference between the two (Wilcoxon signed-rank test: $P < 0.001$).

Replicating the finding of anti-compression in the presence of noise ($s \gg 0$), we found the best-fitting estimates of k in both tasks to be larger than 1 (Fig. 2b; see Supplementary Information for details) for both sample formats (digits and dots: all $k > 1.5$, Wilcoxon signed-rank tests: all $P < 0.05$, uncorrected). More importantly, in the multi-choice task, the estimates of k were significantly larger (mean: 3.40 ± 0.48) compared with the standard task (mean: 1.60 ± 0.20; 2×2 repeated-measures analysis of variance: main effect of task $F_{1,20} = 12.92$, $P = 0.002$). In other words, as we increased integration noise, the observed anti-compression also increased. We also again found evidence for a positive offset bias (see Supplementary Information for details), indicating that participants especially overweighted larger numbers (Supplementary Fig. 2b). Lastly, the analysis revealed no significant differences in any of the above effects between digits and dots displays (all $F_{1,20} < 3.88$, all $P > 0.05$), confirming that selective number integration occurred independent of presentation format (symbolic versus non-symbolic, see above for similar results for visual versus auditory). Together, across all study conditions, humans adopted a non-linear sampling policy that drove accuracy near to the model-predicted maximum, given their estimated noise level and bias (Fig. 2a,b).

To explore these effects at the neural level, we recorded electroencephalograms (EEG) while participants performed the first experiment (Fig. 1a). All sequential samples were fully statistically independent, allowing us to analyse neural responses evoked by individual numbers in the stream (see Methods). Consistent with previous research[13,14], we observed differences in the centro-parietal positivity (CPP) response following the onset of each number in the periods 290–700 ms (visual) and 500–800 ms (auditory) post-onset (Fig. 3a; all time bins $P < 0.01$, false discovery rate corrected). In the visual modality (Fig. 3b, left panel), the CPP response was larger for number 6, reduced for sample 5 and smallest for all other numbers, as indicated by *post hoc* tests (Wilcoxon signed-rank tests: 6 versus 5, $P = 0.008$; 5 versus 4, $P < 0.001$; whereas 4 versus 3, 3 versus 2 and 2 versus 1, all $P > 0.70$; Bonferroni corrected). The auditory condition followed a similar pattern (Fig. 3b, right panel), albeit with noisier and lower-amplitude CPP effects (6 versus 5, $P = 0.28$; 5 versus 4, $P = 0.004$; 4 versus 3, 3 versus 2 and 2 versus 1, all $P > 0.40$; corrected). We fitted the neural amplitude modulations with the absolute decision weights obtained from the non-linear model (purple dots in Fig. 1d), which provided a better fit than the linear model for both auditory and visual conditions (both $P < 0.02$, Wilcoxon signed-rank test on regression deviances). Expanding the CPP analysis to encompass sample order, we observed no interactions between order and number (both $F < 1.4$, both $P > 0.20$), suggesting that the overweighting of larger numbers was invariant across sample positions (1–10).

Next, we employed a multivariate approach (representational similarity analysis (RSA))[6,15] to probe the neural encoding of numbers in more detail and link the neural representations back to categorization behaviour. We computed for each post-sample time point the representational (dis)similarity in EEG signals for each number from 1 to 6 in each of the two categories (12×12 representational

dissimilarity matrix (RDM), based on the Mahalanobis distance; see Methods). We then compared this with predicted RDMs that were created under the assumption that neural distance depended on: (1) the physical properties of the digit, (2) category membership (for example, red versus green font), (3) parity (odd versus even), (4) numerical distance (that is, the pairwise numerical difference between any two numbers, independent of category), and (5) category-dependent numerical distance (see Supplementary Fig. 3a for details). We used recursive orthogonalization (see Methods) to ensure that each model RDM explained unique variance in the observed neural RDM from human subjects.

Figure 4a shows a plot of the time course of correlations (Kendall's tau) between the five model RDMs and the human RDM for each subject in the visual condition. The neural patterns were dominated by a category-independent numerical distance effect (Fig. 4a, purple) that was significant from approximately 200–700 ms after sample onset ($P_{cluster} < 0.001$; cluster-based permutation test); this can also be seen in the grand mean EEG RDM (Fig. 4b). However, additional effects of category (font colour), parity and a category-specific numerical distance effect (distance × category) were also observed (all $P_{cluster} < 0.01$), with the category-specific numerical distance effect peaking late (Fig. 4a, green), consistent with a response-mapped representation. To further visualize these effects, we reduced the dimensionality of the dissimilarity matrix via multidimensional scaling. Visualizing the first three dimensions showed clear effects of numerical distance (x dimension), category (z dimension) and parity (y dimension) (Fig. 4c). Visual inspection of the grand mean EEG RDM (Fig. 4b) may suggest that a numerical distance effect might have arisen mostly by dissimilarity of the numbers 6 and 5 compared with the remaining numbers (1–4; see also the CPP analysis, Fig. 3). Interestingly, however, we found a statistically significant effect even when restricting the analysis to numbers 1–4 (324–574 ms, $P_{cluster} < 0.001$). In other words, multivariate RSA disclosed aspects of a number line representation that were invisible to conventional parietal evoked signals (Fig. 3; see also Supplementary Fig. 3c). However, we observed no systematic EEG–RSA effects in the auditory condition (Supplementary Fig. 3b).

Having established a numerical distance effect in the multivariate EEG patterns, we investigated whether the neural data predicted the distortions in numerical coding observed in the behavioural weighting profiles (Fig. 1d). To test this, we estimated the best-fitting 'neurometric' mapping function predicted from the EEG–RSA patterns by generating model RDMs from hypothetical mapping functions parameterized by k and b (Fig. 4d). We exhaustively searched over values of k (0.4 to 4) and b (−1 to 1) and correlated the predicted model RDMs (both for distance and distance × category effects) with the EEG–RSA pattern in each participant. The best-fitting parametrizations (in terms of maximum mean Kendall's tau correlation) were characterized by values of $k > 1$ (mean: 2.52 ± 0.20, Wilcoxon signed-rank test: $P < 0.001$) and $b > 0$ (mean: 0.15 ± 0.05, $P = 0.005$) (Fig. 4e). In other words, the neurometric number mapping inferred from the EEG–RSA mirrored both key aspects of the psychometric mapping inferred from choice behaviour: (1) exponential overweighting of outlying samples (that is, anti-compression) and (2) an overall weighting bias towards large numbers. Together, these results show strong correspondence of model simulations, choice behaviour and sample-level neural representations in demonstrating 'optimal irrationality'[2] or 'rational inattention'[16] in the presence of noise during sequential information integration.

The present findings build on recent work in which participants compared the average magnitude in two streams of visual items occurring in parallel (for example, side by side on a screen). In this setting, they tend to ignore or downweight the locally weaker of the two simultaneously occurring samples[1] and this behaviour can similarly be accounted for with a selective weighting policy that systematically discards decision information. Given a selective weighting policy, it is possible to construct equally valued streams A, B and C such that participants will systematically choose A > B, B > C and C > A; that is, they will show a classic violation of the rational axiom of transitivity[2]. Nevertheless, in both the present study and the aforementioned one, selective weighting maximized accuracy—that is, it was rational—if one assumes that noise corrupts information integration[2]. We note that the selective weighting policy observed in our experiments tended to overweight larger numbers (for example, 6) rather than all outliers (for example, 1 and 6), as would be predicted by the rational policy under late noise. We leave it to future research to determine whether the offset bias that was observed in both experiments (although not in multi-choice conditions) depends in part on the framing of the task. Finally, recent work has used the selective weighting framework to provide a normative account of the 'robust averaging' (that is, downweighting of outliers, not inliers) of decision information that occurs when stimulus feature values are distributed in an approximately Gaussian fashion across the experiment[17]. In all of these cases, humans seem to have evolved policies that discard information to increase the robustness of decisions in the face of noise corrupting the neural computations associated with information integration.

Behavioural signatures of decision weighting were also reflected in neural signals. The CPP response is an evoked centro-parietal potential that has previously been shown to build up during information integration with an amplitude that reflects the strength of the available decision information[13,14]. Here, we observed a relatively larger CPP response for numbers 5 and 6 in both the auditory and visual domains. The CPP response is most likely related to the well-described P300 potential[18–20] and it may relate here to the detection of the information that is being used to form a decision[21] or to evaluation processes that unfold at the level of each individual sample[19]. The effects were discernable, but considerably more noisy, in the auditory domain compared with the visual domain. This was probably related to unavoidable time-varying differences in the specific physical input associated with each speech sample. Together, these findings offer independent corroborating evidence for the strategy of selective overweighting that we observed in participants' behaviour.

The RSA results revealed a neural representation of an ordered 'number line' for numeric visual symbols. Similar representations of numerical magnitude have previously been reported for non-symbolic numbers; for example, dot displays[22–24], but not for number symbols and digits[22–26]. It is striking that neural patterns recorded at the scalp encode numerical magnitude even when other potentially correlated factors have been accounted for (for example, visual similarity in digits themselves), and future researchers may wish to harness this finding to reveal other aspects of human numerical cognition. Here, however, we emphasize that in the visual domain, the neural representation of the number line was distorted in exactly the way predicted by behavioural data. Interestingly, although it is theoretically possible that the CPP response could account for the pattern similarity effects revealed with RSA, it seems unlikely that this is the case in our data. For example, we found that participants' 'number line' in decision weighting explained RSA variance not accounted for by the CPP response (Supplementary Fig. 3c). In addition, we observed no RSA effects for the auditory condition despite statistically reliable differences in the CPP response. The reasons for this discrepancy are unclear. It could be related to the difficulty in establishing high-precision neural patterns in time-locked data for time-varying speech items, or it might reflect computational differences in the processing of auditory stimuli.

Methods

Participants.. Healthy volunteers ($n = 24$; 12 females and 12 males; age: 26.6 ± 2.8) participated in the experiment after providing written informed consent. The study was approved by the ethics commission of the Free University Berlin and was conducted in accordance with the Human Subjects Guidelines of the Declaration of Helsinki.

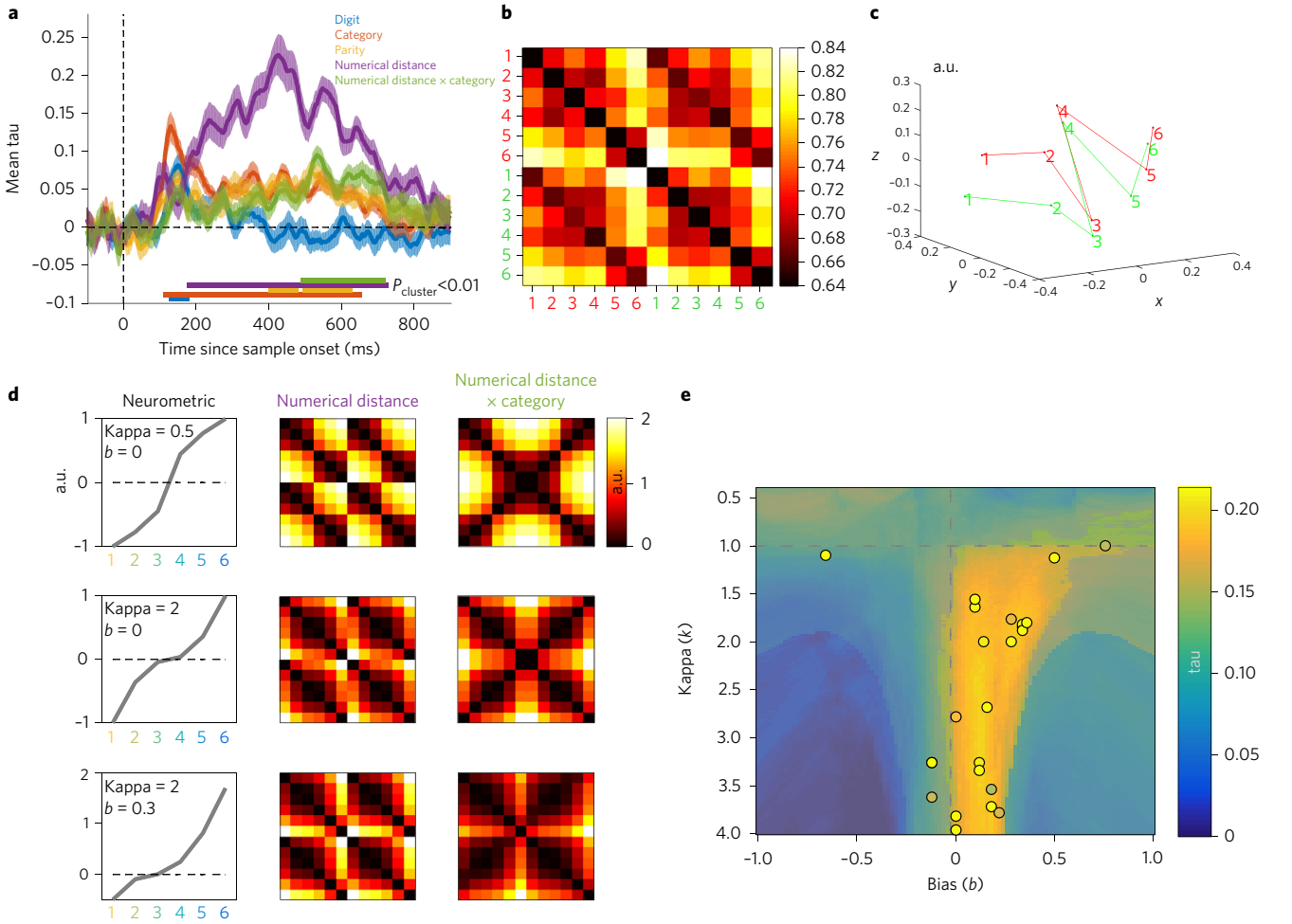

Figure 4 | Representational similarity analysis. a, Time course of correlations (Kendall's tau) between orthogonalized model RDMs for different sample features (see Supplementary Fig. 3a) and the observed EEG–RSA patterns following each sample in the visual condition ($n = 24$). The coloured shaded regions show s.e.m. The marker lines on the bottom indicate significant differences from zero. For the auditory results, see Supplementary Fig. 3b. **b**, Grand-mean EEG–RDM for a representative time window (200–600 ms) in the visual condition. **c**, Three-dimensional illustration of the first three dimensions (plotted on the *x*, *y* and *z* axes) of a multidimensional scaling of the EEG–RDM shown in **b**. Red and green denote the sample colours (see Fig. 1a). **d**,**e**, Neurometric mapping functions estimated from EEG–RSA. **d**, Middle and right: unidimensional model RDMs predicted under different parametrizations of hypothetical neurometric mapping functions (illustrated in left; *y* axis: hypothetical weight in neural encoding (a.u.)). **e**, Grand mean correlations (collapsed over distance and distance × category RDMs from **d**) between model RDMs and observed EEG–RDMs over values of kappa *k* and bias *b*. The dashed grey lines delineate the parameter space for $b = 0$ (symmetric mapping) and $k = 1$ (linear mapping). Maximum mean correlations were observed for values of $b > 0$ and $k > 1$ (see Results). The transparent mask highlights the parameter space of significant positive correlation with both the distance and the distance × category RDM from **d** (Wilcoxon signed-rank tests; all pixels $P_{conjunction} < 0.001$, uncorrected and exceeding the symmetric linear model). The dots show the maximum mean tau for each participant (some dots are covered by others).

Stimuli, task and procedure. On each trial, ten numbers ('samples') were presented in sequence at a rate of 350 ms (Fig. 1a). In the visual condition, digits (font: Arial, approximate visual angle: 1.8°) were presented at fixation in either a green or red font for 280 ms, followed by a 70 ms blank period. In the auditory condition, German number words were played with either a female or male voice. Speech samples were taken from a public repository (http://www.freesound.org), time-compressed to a common length of 350 ms (using the PSOLA algorithm in Adobe Audition CC; http://www.adobe.com) and loudness normalized. Each sample was independently and randomly drawn with uniform probability from a pool of 12 possible items consisting of the numbers 1–6 in each of the two categories (that is, the number of samples drawn from each category was fully randomized). Following the offset of the final item in the sequence, participants were given 2 s to indicate by key press (left or right hand, counter-balanced between participants) which of the two sample categories contained the higher average numerical value. Median response times averaged 487 ms (visual) and 484 ms (auditory). After 100 ms, correct responses were rewarded with a bell ('bling') sound, whereas errors were fed back by a 'buzz' sound. After a brief wait period (500–1500 ms, randomly varied), the next trial started. The onset of a trial (500 ms before sequence onset), as well as the response periods (350 ms after the onset of the last item in a sequence), were signalled by a small central fixation point briefly changing in colour (between grey and white). Participants were instructed to maintain fixation throughout all trials (including the auditory condition), and this was aided by a head support (SR Research) to avoid movements. After several practice runs, each participant performed six blocks of 100 trials (three in each modality condition, in alternating order), providing 3,000 sample presentations per modality and participant.

Psychophysical model. In our simulations, for convenience, we defined *X* as ranging between −1 and 1 in six equidistant steps, corresponding to the six numerical magnitudes (1–6) used in the experiment. We characterized the mapping of sample information *X* onto a subjective decision value (*dv*) as a family of (sign-preserving) exponential functions:

$$dv = \frac{X+b}{|X+b|} \times |X+b|^k \quad (1)$$

where $k < 1$ implies a relative downweighting and $k > 1$ a relative upweighting of outlying samples (Fig. 1b, left panel). The special case where $k = 1$ corresponds to a linear mapping; that is, $dv = X$. Parameter *b* accounts for a potential asymmetric weighting bias, in terms of an offset of the mapping function relative to its

indifference point ($dv = 0$). A value of $b = 0$ corresponds to a point symmetric mapping, whereas $b \neq 0$ implies an up- or downward-shifted asymmetric function (Fig. 1b, right panel).

Our initial goal was to evaluate model performance across different values of k and b. Here, it is important to consider that different parameterizations of the mapping function in equation (1) differ in the absolute decision value that is obtained by transformation of perceptual inputs, X. This absolute decision value is, in turn, related to the probability of a correct choice being made (see equation (4)). Assuming that decisional gain is a limited cognitive resource (that is, a quantity that should not change between the models we test; for example, reflecting an upper limit on the number of spikes produced by the relevant neurons), we computed for each transformation (equation (1)) its multiplicative gain factor, g:

$$g = \frac{\sum |f + b|^k}{\sum |f|} \tag{2}$$

which quantifies the extent to which a feature space, f (here, the six equiprobable values of X), is transformed into a dv space whose absolute values are larger (or smaller) than the absolute values in f. Using g as a normalization factor, the trial-level decision value, DV, is given by the sum over the 10 samples of each sequence (Fig. 1a):

$$\mathrm{DV} = \sum_{i=1}^{10} \frac{dv_i \times c_i}{g} \tag{3a}$$

where c_i is a dummy variable that encodes the category of a sample (for example, $c_{red} = 1$; $c_{green} = -1$). To additionally model a potential leak, l, of decision value over time (Supplementary Fig. 1b), we extended equation (3a) by a simple exponential function over samples, i (ref. [27]):

$$\mathrm{DV} = \sum_{i=1}^{10} \frac{dv_i \times c_i}{g} \times l^{10-i} \tag{3b}$$

Lastly, the trial-level DV was transformed into a choice probability, CP, using a logistic choice function with noise term sigma (s):

$$\mathrm{CP} = \frac{1}{1 + e^{\frac{-\mathrm{DV}}{s}}} \tag{4}$$

We refer to s as 'late' or integration noise, denoting noise that occurs at processing stages downstream from perceptual sample encoding. Such noise could arise during integration or at the response itself, but we note that the compact parametrization of late noise in equation (4) is equivalent to adding a (constant) noise term to each dv_i.

To simulate model accuracy (Fig. 1c), model choices were generated by randomly drawing from a binomial distribution with a binomial probability, p, equal to CP, where CP was computed trial by trial according to equations (1–4). When fitting human choice data, we included a constant term to account for potential motor biases (for example, towards left versus right responses). To avoid parameter instabilities, we fitted the model without gain normalization and rescaled s by dividing it by g, which warrants formal equivalence to equations (1–4). Parameter estimates were obtained by minimizing the negative log-likelihood of the model given each participant's single-trial responses across values of k (0.1 to 10), b (−1 to 1), s (0.01 to 8, unnormalized) and (in models with leakage) l (0 to 1). In two participants in the visual condition, the model without leakage (equation (3a)) yielded exceedingly large raw estimates of s. However, the group-level results were robust to either inclusion or exclusion of these participants. Quantitative model comparisons (for example, between exponential and linear models) were corrected for model complexity based on the AIC. To evaluate model performance against human choice behaviour, we again generated binomial model choices (simulations), but this time using the individual best-fitting model parametrizations and the same sample sequences as those presented in the human experiment. We then compared the choice data of human and model observers using conventional statistical analyses. Choice probabilities associated with each sample number (1–6) were inferred from the relative frequency of choosing a sample's category (that is, its colour or speaker) at the end of a trial and were transformed into estimates of (signed) decision weight with an indifference point at zero (that is, decision weight = choice probability − 0.5; Fig. 1d, dual y axes). Evaluation against model predictions was complemented by model-free tests for symmetry (comparing the absolute decision weights of 1 versus 6, 2 versus 5 and 3 versus 4) and linearity (an omnibus test of linear regression residuals across numbers 1–6) of the human weighting functions.

EEG recording and analysis. We recorded 64-channel EEGs (BioSemi ActiveTwo) configured according to the extended 10–20 system. Ocular activity was recorded via adhesive electrodes placed in bipolar montages around the eyes (horizontal and vertical) and was additionally monitored using an EyeLink 1000 camera (SR Research). EEG signals were digitized at 2,048 Hz, off-line referenced to common average, filtered (0.5–45 Hz) and down-sampled to 256 Hz. The EEGs were corrected for eye blinks using adaptive spatial filtering[28] and epoched around each individual sample (−100 to 900 ms relative to sample onset). Bad channels were identified by visual inspection. Residual artefacts were rejected by excluding epochs with amplitudes of greater than 80 μV from the analysis. The artefact-free epochs were baseline subtracted (−100 to 0 ms) and smoothed with a sliding 50 ms Gaussian kernel. EEG analyses were performed in MATLAB (R2016a; MathWorks) using functions from SPM12 (build 6470) for M/EEG (www.fil.ion.ucl.ac.uk/spm/), including, FieldTrip (http://www.ru.nl/neuroimaging/fieldtrip) and custom MATLAB code.

CPP responses. All EEG analyses were performed on the individual sample level. In each participant and modality condition, the mean waveform (averaged over all samples) was subtracted from each individual epoch, effectively eliminating the stimulus-onset response to the current and subsequent samples (note that epochs overlapped with up to two subsequent sample onsets; Fig. 1a). Epochs of the same sample type were averaged and subjected to conventional statistical analysis. Based on previous CPP response findings[14,20,29], signals were pooled over centro-parietal channels (CP1, P1, POz, Pz, CPz, CP2 and P2). Time windows for CPP analysis were identified using a leave-one-out procedure to preclude circular inference. For each participant, CPP amplitudes were averaged over the adjacent significant time bins ($P < 0.01$, false discovery rate corrected) that exhibited the strongest overall amplitude modulations in the remaining 23 participants based on non-parametric omnibus tests over sample types.

Representational similarity analysis. The pre-processed channel data were projected onto principal components retaining 99% of the variance[30]. Multivariate (dis)similarity was assessed in terms of the pairwise Mahalanobis distance between the mean-subtracted component patterns associated with each sample type (numbers 1–6 per sample category, yielding a 12 × 12 RDM at each time point), using the residual single-trial variance at each time point for noise normalization[31].

To test the extent to which sample information was encoded in the time-course of the EEG–RDM, we created hypothetical model RDMs for the following features of interest (Supplementary Fig. 3a): (1) physical number, with minimum dissimilarity between identical numbers and maximum dissimilarity between all other pairs, regardless of sample category; (2) sample category, with minimum and maximum dissimilarity between same and different category samples; (3) numerical parity, with minimum and maximum dissimilarity between numbers of the same and different parity (even or uneven); (4) numerical distance, with dissimilarity linearly increasing as a function of the numerical difference between any two numbers, independent of sample category; and (5) category-dependent numerical distance, where the encoding of numerical distance within each sample category is the same as in (4), but is inverted between the two categories (in terms of a numerical distance × category interaction—that is, a 6 in one category is predicted to be similar to a 1 in the other category). The latter is expected to occur if numerical value representations were response mapped; that is, if they systematically differed in driving left (for example, 'red') versus right (for example, 'green') key choices.

Each model RDM was recursively orthogonalized with respect to all other model RDMs using the Gram–Schmidt process (Supplementary Fig. 3a). Then, each model RDM was correlated with the EEG–RDM at each peri-sample time point using Kendall's tau correlation coefficient. All correlations were computed over the upper RDM triangle, excluding all redundant elements and the diagonal. Significant correlations were identified using cluster-based permutation tests[32] over time points (1,000 iterations, cluster-defining threshold $P < 0.01$, Wilcoxon signed-rank tests, uncorrected). Subsequent analyses were performed on the mean EEG–RDM in a representative time window (200–600 ms). For dimensionality-reduced visualization, we used classical multidimensional scaling as implemented in MATLAB, selecting the dimensions with the largest three eigenvalues in explaining the grand mean RDM.

Supporting experiment. Methods and additional analyses for the supporting experiment are presented in the Supplementary Information.

Statistical analyses. Behavioural and modelling results were analysed using non-parametric tests (two-sided) as detailed in the Methods and Results. Time windows of significant effects in neural data were identified using leave-one-out and permutation procedures, as explained in the Methods (CPP and RSA analyses). Complementary analysis of variance results were based on Greenhouse–Geisser corrected degrees of freedom where appropriate.

Code availability. Custom code used in the analysis has been made available at https://github.com/summerfieldlab/Spitzer_etal_2017.

Data availability. The data that support the findings of this study are available at https://github.com/summerfieldlab/Spitzer_etal_2017. Raw EEG files are available from the corresponding author on request.

Received 23 May 2017; accepted 13 June 2017; published 17 July 2017

References

1. Tsetsos, K., Chater, N. & Usher, M. Salience driven value integration explains decision biases and preference reversal. *Proc. Natl Acad. Sci. USA* **109,** 9659–9664 (2012).
2. Tsetsos, K. *et al.* Economic irrationality is optimal during noisy decision making. *Proc. Natl Acad. Sci. USA* **113,** 3102–3107 (2016).
3. Brezis, N., Bronfman, Z. Z., Jacoby, N., Lavidor, M. & Usher, M. Transcranial direct current stimulation over the parietal cortex improves approximate numerical averaging. *J. Cogn. Neurosci.* **28,** 1700–1713 (2016).
4. Brezis, N., Bronfman, Z. Z. & Usher, M. Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Sci. Rep.* **5,** 10415 (2015).
5. Malmi, R. A. & Samson, D. J. Intuitive averaging of categorized numerical stimuli. *J. Verbal Learning Verbal Behav.* **22,** 547–559 (1983).
6. Kriegeskorte, N. & Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* **17,** 401–412 (2013).
7. Scott, B. B., Constantinople, C. M., Erlich, J. C., Tank, D. W. & Brody, C. D. Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats. *eLife* **4,** e11308 (2015).
8. Wyart, V. & Koechlin, E. Choice variability and suboptimality in uncertain environments. *Curr. Opin. Behav. Sci.* **11,** 109–115 (2016).
9. Gibbon, J. Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* **84,** 279–325 (1977).
10. Moyer, R. S. & Landauer, T. K. Time required for judgements of numerical inequality. *Nature* **215,** 1519–1520 (1967).
11. Dehaene, S., Dupoux, E. & Mehler, J. Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* **16,** 626–641 (1990).
12. Van Opstal, F., de Lange, F. P. & Dehaene, S. Rapid parallel semantic processing of numbers without awareness. *Cognition* **120,** 136–147 (2011).
13. Kelly, S. P. & O'Connell, R. G. Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J. Neurosci.* **33,** 19434–19441 (2013).
14. O'Connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* **15,** 1729–1735 (2012).
15. Nili, H. *et al.* A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10,** e1003553 (2014).
16. Woodford, M. Prospect theory as efficient perceptual distortion. *Am. Econ. Rev.* **102,** 41–46 (2012).
17. Li, V. L., Castañón, S. H., Solomon, J. A., Vandormael, H. & Summerfield, C. Robust averaging protects decisions from noise in neural computations. Preprint at *bioRxiv* https://doi.org/10.1101/147744 (2017).
18. Twomey, D. M., Murphy, P. R., Kelly, S. P. & O'Connell, R. G. The classic P300 encodes a build-to-threshold decision variable. *Eur. J. Neurosci.* **42,** 1636–1643 (2015).
19. Donchin, E. & Coles, M. G. H. Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* **11,** 357–374 (1988).
20. Sutton, S., Braren, M., Zubin, J. & John, E. R. Evoked-potential correlates of stimulus uncertainty. *Science* **150,** 1187–1188 (1965).
21. Picton, T. W. The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* **9,** 456–479 (1992).
22. Bulthé, J., De Smedt, B. & Op de Beeck, H. P. Visual number beats abstract numerical magnitude: format-dependent representation of Arabic digits and dot patterns in human parietal cortex. *J. Cogn. Neurosci.* **27,** 1376–1387 (2015).
23. Eger, E. *et al.* Deciphering cortical number coding from human brain activity patterns. *Curr. Biol.* **19,** 1608–1615 (2009).
24. Harvey, B. M., Klein, B. P., Petridou, N. & Dumoulin, S. O. Topographic representation of numerosity in the human parietal cortex. *Science* **341,** 1123–1126 (2013).
25. Dehaene, S. The organization of brain activations in number comparison: event-related potentials and the additive-factors method. *J. Cogn. Neurosci.* **8,** 47–68 (1996).
26. Libertus, M. E., Woldorff, M. G. & Brannon, E. M. Electrophysiological evidence for notation independence in numerical processing. *Behav. Brain Funct.* **3,** 1 (2007).
27. Wyart, V., Myers, N. E. & Summerfield, C. Neural mechanisms of human perceptual choice under focused and divided attention. *J. Neurosci.* **35,** 3485–3498 (2015).
28. Ille, N., Berg, P. & Scherg, M. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J. Clin. Neurophysiol.* **19,** 113–124 (2002).
29. Spitzer, B., Blankenburg, F. & Summerfield, C. Rhythmic gain control during supramodal integration of approximate number. *NeuroImage* **129,** 470–479 (2016).
30. Grootswagers, T., Wardle, S. G. & Carlson, T. A. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* **29,** 677–697 (2017).
31. Nili, H., Walther, A., Alink, A. & Kriegeskorte, N. Inferring exemplar discriminability in brain representations. Preprint at *bioRxiv* https://doi.org/10.1101/080580 (2016).
32. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164,** 177–190 (2007).

Acknowledgements

This work was supported by grants from the German Research Foundation to B.S. (DFG SP 1510/1-1 and DFG SP 1510/2-1) and a European Research Council Starter Grant (281628) to C.S. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank V. Li, T. Flesch and H. Nili for helpful suggestions and scripts, F. Blankenburg for resources, A. Epure for help with data acquisition and R. Kievit for helpful comments on a previous version of the manuscript.

Author contributions

B.S. designed the experiments with contributions from L.W. and C.S. L.W. and B.S. conducted the experiments. B.S. and C.S. developed the analysis approach. B.S. analysed the data with contributions from C.S. B.S. and C.S. wrote the paper.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

How to cite this article: Spitzer, B., Waschke, L. & Summerfield, C. Selective overweighting of larger magnitudes during noisy numerical comparison. *Nat. Hum. Behav.* **1,** 0145 (2017).

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

Competing interests

The authors declare no competing interests.