

EEG-representational geometries and psychometric distortions in approximate numerical judgment

Stefan Appelhoff^{1,2*}, Ralph Hertwig¹, Bernhard Spitzer^{1,2}

¹ Center for Adaptive Rationality, Max Planck Institute for Human Development

² Research Group Adaptive Memory and Decision Making, Max Planck Institute for Human Development

*corresponding author; email: appelhoff@mpib-berlin.mpg.de; tel: +49 30 82406-496; address: Lentzeallee 94, 14195 Berlin, Germany

Running title: EEG-representational geometries and psychometric distortions

Abstract

When judging the average value of sample stimuli (e.g., numbers) people tend to either over- or underweight extreme sample values, depending on task context. In a context of overweighting, recent work has shown that extreme sample values were overly represented also in neural signals, in terms of an anti-compressed geometry of number samples in multivariate electroencephalography (EEG) patterns. Here, we asked whether neural representational geometries may also reflect underweighting of extreme values (i.e., compression) which has been observed behaviorally in a great variety of tasks. We used a simple experimental manipulation (instructions to average a single-stream or to compare dual-streams of samples) to induce compression or anti-compression in behavior when participants judged rapid number sequences. Model-based representational similarity analysis (RSA) replicated the previous finding of neural anti-compression in the dual-stream task, but failed to provide evidence for neural compression in the single-stream task, despite the evidence for compression in behavior. Instead, the results suggested enhanced neural processing of extreme values in either task, regardless of whether extremes were over- or underweighted in subsequent behavioral choice. We further observed more general differences in the neural representation of the sample information between the two tasks. The results suggest enhanced processing of extreme values as the brain's default. Such a default raises new questions about the origin of common psychometric distortions, such as diminishing sensitivity for larger values.

Keywords: decision making, numerical cognition, electroencephalography, MVPA, computational modeling

Introduction

When making decisions about magnitudes such as numbers, people tend to distort sample information away from its true value. A commonly observed distortion is a compression of magnitude, where extreme (or outlying) samples receive relatively less weight than prescribed by a linear and, according to common interpretation, normative transformation of objective into psychological or subjective values (Bernoulli, 1954; Fechner, 1860; Juechems et al., 2021; Li et al., 2017; Tversky & Kahneman, 1992; Vandormael et al., 2017). However, in some task contexts, the opposite type of distortion has been observed, that is, an anti-compression, where extreme or outlying samples are overweighted (Clarmann von Clarenau et al., 2022; Kunar et al., 2017; Luyckx et al., 2019; Spitzer et al., 2017; Tsetsos et al., 2012; Vanunu et al., 2020).

Using electroencephalographic (EEG) recordings and multivariate representational similarity analysis (RSA), recent work has identified a potential neural signature of such psychometric distortions. During processing of symbolic number samples, multivariate EEG patterns have been characterized by a “numerical distance effect”, where the representational similarity of, for instance, “4” and “5” is larger than that between “4” and “6”, which in turn is larger than that between “3” and “7”, and so forth (Appelhoff et al., 2022; Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). Intriguingly, in a multi-sample decision task that promoted anti-compression of number samples in behavior, the “neural numberline” underlying the numerical distance effect was found to be anti-compressed as well (Appelhoff et al., 2022; Luyckx et al., 2019; Spitzer et al., 2017). The findings suggested that behaviorally relevant distortions in multi-sample decisions may occur already when the individual samples are being processed.

However, such “neurometric” signature of psychometric distortion has thus far only been reported in tasks that promoted anti-compression, that is, a selective overweighting of extreme (outlying) sample values in behavior. A much more common observation in other task contexts is a *compression* of magnitude, where extreme values are underweighted, for instance, in psychophysical tasks (Fechner, 1860; Stevens, 1957; Wyart et al., 2012), in studies of numerical cognition (Dehaene, 2003; Longo & Lourenco, 2007; Nieder & Dehaene, 2009; Nieder & Miller, 2003), and in behavioral economics experiments (Kellen et al., 2016; McAllister & Tarbert, 1999; Tversky & Kahneman, 1992). To what extent large-scale neural patterns, as recorded with human EEG, may also reflect psychometric compression is still unknown.

In the present study, we capitalized on recent progress in understanding the experimental factors that may mediate whether people compress or anti-compress magnitudes in decision making (see also Summerfield & Parpart, 2022). Specifically, in a recent behavioral study, we found compression when judging the average of a single stream, but anti-compression when comparing dual streams of number samples (Clarmann von Clarenau et al., 2022). Here, we adopted this experimental manipulation to examine the neural signatures of compressive (as compared to anti-compressive) number processing in multivariate EEG patterns.

Behaviorally, the results confirmed a compression of numerical values in the single-stream task, and an anti-compression of the same values in the dual-stream task. In neural signals, we replicated the finding of an anti-compressed number representation in the dual-stream task. Surprisingly, however, we found no evidence for neural compression in the single-stream task. Instead, we observed more general differences in the neural representation of the sample information. Whereas in the dual-stream task, the samples' neural geometry predominantly reflected their abstract magnitude, the single-stream task was associated with a more direct, non-quantitative representation of the concrete sample stimuli. The results relativize the diagnosticity of sample-level EEG-metrics for psychometric distortions. They also suggest a default mode of processing, namely, enhanced neural processing of extreme sample values, regardless of whether they are over- or underweighted in subsequent behavior.

Results

Participants (n=30) performed two different variants of a sequential integration task where they observed sequences of ten digits (ranging between 1 and 9, colored in red or blue; Fig. 1a). In the single-stream variant (“averaging” task), participants were asked to report whether the average of all ten number samples (regardless of color) was larger or smaller than 5. In the dual-stream variant (“comparison” task), they were asked to indicate whether the red or the blue samples had the higher average value.

Behavioral results

As expected, mean choice accuracy (Fig. 1b) was higher in the single-stream task ($80.5\% \pm 0.8\%$ SE) than in the dual-stream task [$76.3\% \pm 0.7\%$ SE; $t(29)=5.51$, $p<0.001$, $d=1.02$; paired t-test]. This suggests that comparing two streams was more difficult than averaging a single-stream (of otherwise physically identical inputs; Fig. 1a).

To characterize how the numerical value (1-9) of a sample influenced subsequent choice, we calculated model-free decision weights (see *Materials and Methods*). Descriptively, the weighting curve showed a concave shape (indicating compression) in the single-stream task (Fig. 1c, *left*), whereas an convex shape (indicating anti-compression) was evident in the dual-stream task (Fig. 1c, *right*).

For quantitative analysis, we fitted a psychometric model (see *Materials and Methods*), which characterizes the transformation of sample values (1-9) as a sign-preserving power function with exponent kappa (k ; where $k<1$ indicates compression and $k>1$ anti-compression). The model further includes parameters for overall bias (b) towards larger or smaller numbers ($b > 0$ or $b < 0$) and decision noise (s , see *Materials and Methods*).

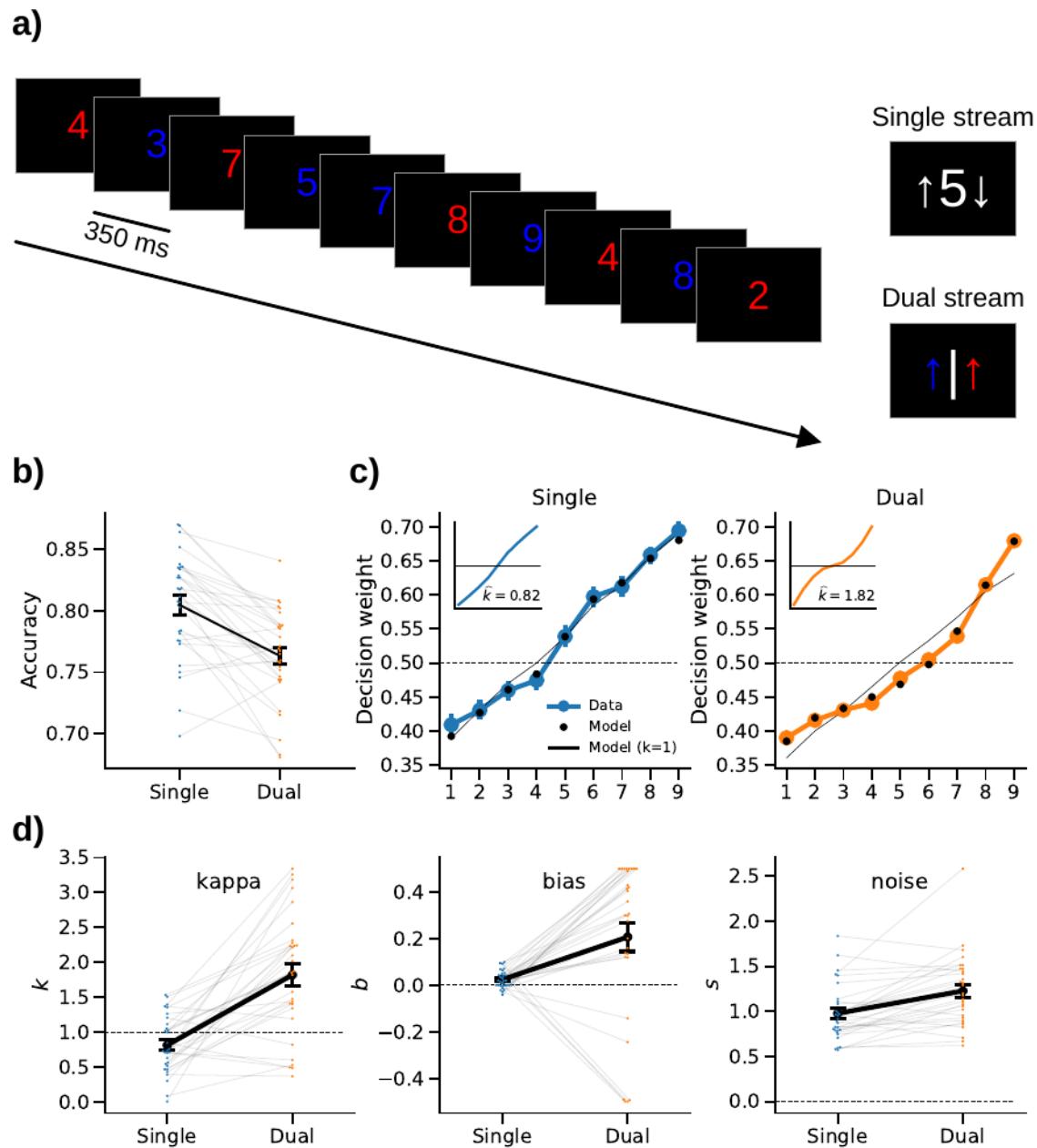


Figure 1. Experimental paradigm and behavioral results. a) Each participant performed two variants of a sequential number integration task. *Left*: In both variants, participants viewed a stream of ten rapidly presented digits (five in red and five in blue color; in random serial order) drawn from 1 to 9 (uniform random). In the single-stream averaging task, participants were asked to judge whether the average value of all ten samples was higher or lower than 5 (ignoring the samples' colors). In the dual-stream comparison task, participants were asked to report whether the red or the blue samples had the higher average value. *Right*: In both tasks, the response mapping onto left/right button presses was randomized across trials and cued only after sample presentation, in order to avoid motor preparation confounds. **b)** Mean accuracy (proportion correct choices) in the two tasks. **c)** Decision weights of number values in the single-stream (left) and dual-stream (right) tasks. Inset plots illustrate the shape of distortion implied by the best-fitting k (see panel **d**) according to model Eq. 1, with b set to 0 for visual comparison between task conditions. **d)**

Parameter estimates from fitting our psychometric model to the empirical choice data (cf. c). Error bars in all panels show SE.

The best-fitting parameter estimates are shown in Fig. 1d. Of main interest was parameter kappa (k), which indicates the extent to which a weighting policy is compressed or anti-compressed relative to a linear weighting ($k = 1$). Indeed, k was significantly smaller than 1 in the single-stream task [$M=0.82$, $SE=0.07$, $t(29)=2.43$, $p=0.02$, $d=0.44$; one-sample t-test against 1] and significantly larger than 1 in the dual-stream task [$M=1.82$, $SE=0.16$, $t(29)=5.21$, $p<0.001$, $d=0.95$], which confirms robust compression in single-stream averaging, and robust anti-compression in dual-stream comparison. The essential difference in transformation (Eq. 1) implied by these k -values is illustrated in Fig. 1c (insets). While the transformation is concave (i.e., shallower towards the extremes) in the single-stream task, it is convex (i.e., steeper towards the extremes) in the dual-stream task.

Examining bias (b), we found significantly positive values both in the single-stream task [$M=0.02$, $SE=0.01$, $t(29)=3.52$, $p=0.001$, $d=0.64$] and in the dual-stream task [$M=0.21$, $SE=0.06$, $t(29)=3.34$, $p=0.002$, $d=0.61$]. Thus, judgments in both tasks were overall biased towards larger numbers, which is consistent with previous work (Clarmann von Clarenau et al., 2022; Luyckx et al., 2019; Spitzer et al., 2017). Finally, noise (s) was significantly higher in the dual-stream ($M=1.23$, $SE=0.07$) than in the single-stream task [$M=0.98$, $SE=0.06$; $t(29)=4.40$, $p<0.001$, $d=0.70$; paired t-test]. This is consistent with the lower level of accuracy in the dual-stream task (see Fig. 1b).

Together, our experimental manipulation was successful in inducing opposite types of psychometric distortions, with identical stimulus inputs in the two task conditions. Whereas decision weighting was compressed (suggesting relative underweighting of extreme values) in the single-stream task, it was anti-compressed (suggesting relative overweighting of extreme values) in the dual-stream task.

EEG results

Multivariate (RSA) results

Turning to the EEG data, we first examined the encoding of sample information in multivariate ERP patterns using RSA (see *Materials and Methods*). Specifically, we examined in each of the two tasks (single-stream and dual-stream) the extent to which RSA patterns encoded (i) the concrete digit that was shown as sample stimulus (e.g., “4” or “8”), (ii) its color (i.e., red or blue), and (iii) the numerical magnitude information in a sample (i.e., 1-9, in terms of a numerical distance effect; see *Materials and Methods* and Fig. 2a).

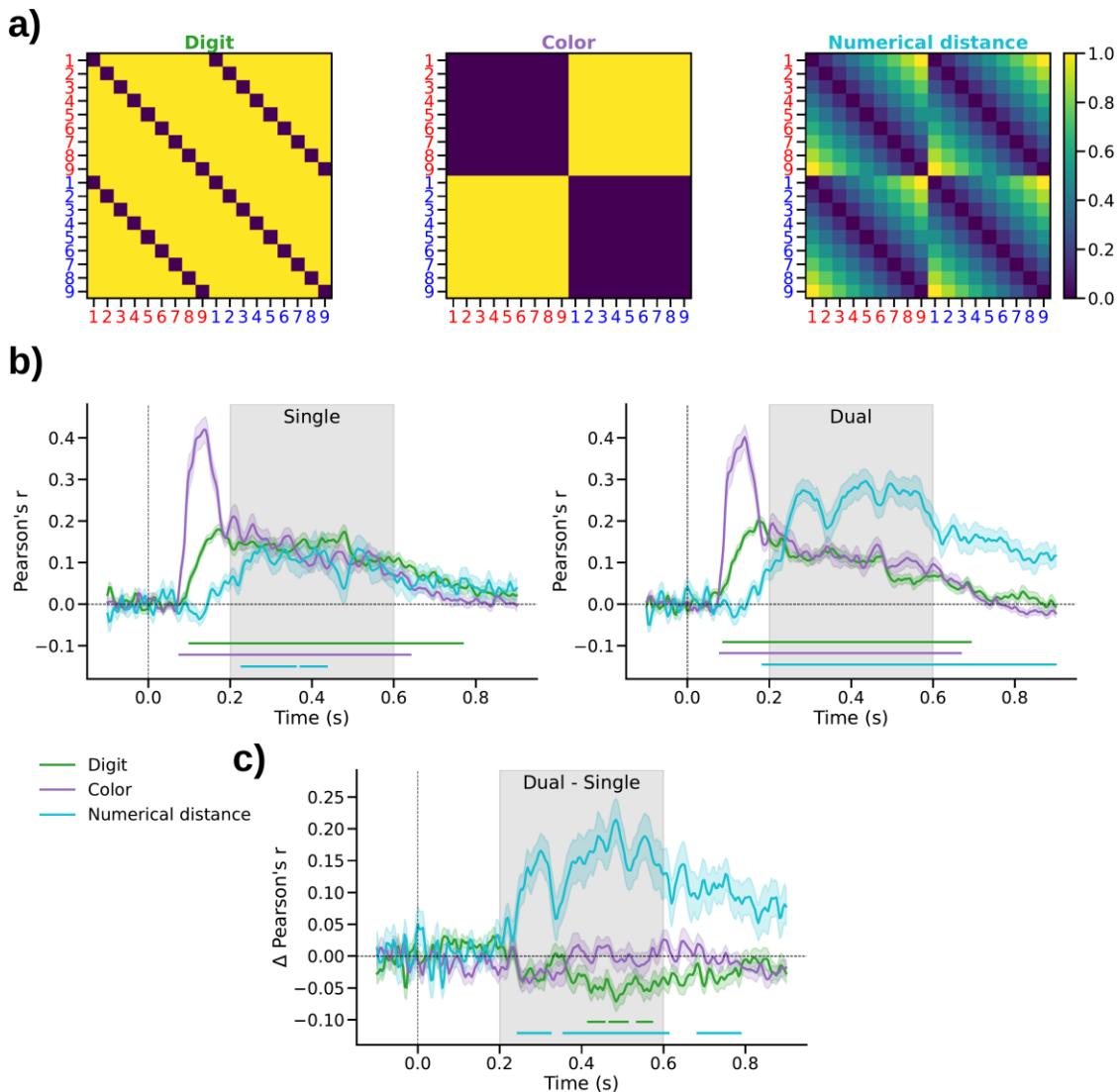


Figure 2. RSA results. **a)** Model RDMs encoding individual sample attributes. *Left*, “Digit” model encoding the unique number symbols; *Middle*, “Color” model encoding whether a sample was red or blue; *Right*, “Numerical distance” model encoding the samples’ magnitude (1-9). For visual clarity, the model RDMs are illustrated before orthogonalization (see *Materials and Methods*). **b)** Correlation between orthogonalized model RDMs and the empirical ERP-RDMs in the single-stream (left) and dual-stream (right) task. Colored shadings show SE. Marker lines on bottom indicate time windows of significant differences from zero. Gray shading outlines the time window used in subsequent neurometric analysis (see Fig. 4 below). **c)** Difference in correlation between the single- and dual-stream tasks. Same conventions as in **b**.

The visual attributes of a sample (i.e., its color and digit shape) were encoded early on, from approximately 100 to 700 ms after sample onset, in both tasks (Fig. 2b all $p_{\text{cluster}} < 0.001$). From approx. 200 ms on, the RSA patterns also encoded the samples’ numerical magnitude, in terms of a significant numerical distance effect (single-stream: $p_{\text{cluster}} < 0.001$; dual-stream: $p_{\text{cluster}} < 0.001$), thus replicating and extending previous work (Appelhoff et al., 2022; Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). Descriptively, the numerical distance

effect observed in the single-stream task, while robustly significant, appeared weaker than that in the dual-stream task.

Comparing the RSA time courses between the two task conditions (Fig. 2c) confirmed that the numerical distance effect was significantly stronger in the dual-stream task ($p_{\text{cluster}} < 0.002$). Surprisingly, we found no difference in color encoding between the two tasks, even though color was task-relevant only in the dual-stream task, but not in the single-stream task. Instead, in the single-stream task, we observed a relatively stronger representation of the concrete digit (i.e., the unique number symbol) that had been displayed ($p_{\text{cluster}} < 0.005$). This effect was evident in a relatively late time window (approximately 400-600 ms, that is, only after the early visual encoding of digits and color).

Together, multivariate ERP patterns in both task conditions robustly encoded information about the sample's color, the number symbol it showed, and its numerical magnitude. The representation of samples in the single-stream task, however, showed qualitative differences, in terms of a relatively weaker encoding of numerical magnitude, and a relatively stronger encoding of the concrete sample stimuli.

Neurometric RSA results

Next, we examined potential distortions of the “neural numberline” underlying the neural magnitude representation disclosed in the above RSA results. To this end, we parameterized the numerical distance model (Fig. 2d, *right*) to reflect distortions by k (compression/ anti-compression) and b (bias towards/against larger numbers), analogously as in our psychometric model (see *Materials and Methods*, Eq. 1). We then used exhaustive gridsearch to determine for each participant the parameter combination with which the model fitted the data best. Fig. 3b illustrates the improvement in fit (in terms of Δr relative to the standard model with $k = 1$ and $b = 0$; cf. Fig. 2d, *right*) in a representative time window (0.2-0.6 s; cf Fig. 2c and 3a). Note that neurometric mapping was performed using a log scale of k (where $\log(k) = 0$ corresponds to $k = 1$, see Fig. 3b) to avoid fitting bias (see *Materials and Methods*).

The results (Fig. 3b) replicated our previous finding of a significant distortion of the neural numberline in the dual-stream task (Fig. 3b *right*). Specifically, like in our previous work (Appelhoff et al., 2022; Spitzer et al., 2017), we observed neurometric estimates of $k > 1$ (i.e., anti-compression) and of $b > 0$ (i.e., a bias towards larger numbers; $p=0.013$, FDR-corrected), which mirrors the pattern observed in the behavioral data (cf. Fig. 1c-d, *orange*). However, contrary to our expectations, we found no evidence for a neurometric compression in the single-stream task (Fig. 3b, *left*), where the psychometric weighting in behavior was clearly compressed (cf. Fig. 1c-d, *blue*). Descriptively, the neurometric map in the single-stream task indicated a pattern similar to that in the dual-stream task (i.e., anti-compression $k > 1$, and positive bias, $b > 0$). However, the improvement in fit over the linear/unbiased model was weak and statistically non-significant (Fig. 3b, *left*), potentially reflecting that the numerical distance effect in the single-stream task was overall weaker (cf. Fig. 2b).

Examining the mean neurometric parameter estimates in the two tasks statistically, they showed significant anti-compression ($k > 1$) both in the dual-stream [$M=3.05$, $SE=0.40$; $t(29)=5.16$, $p<0.001$, $d=0.94$] and in the single-stream task [$M=3.81$, $SE=0.57$; $t(29)=4.90$, $p<0.001$, $d=0.89$, t -tests against 1]. Direct comparison of neurometric k between the two tasks showed no significant difference [$t(29)=1.25$, $p=0.22$, $d=0.28$, paired t -test]. A positive offset bias (b) was evident in the dual-stream task [$M=0.18$, $SE=0.04$; $t(29)=3.97$, $p<0.001$, $d=0.73$] but not in the single-stream task [$M=0.04$, $SE=0.05$; $t(29)=0.75$, $p=0.46$, $d=0.14$, t -tests against 0; difference between tasks $t(29)=2.27$, $p=0.03$, $d=0.54$, paired t -test]. Together, the neurometric RSA results yielded no evidence for a compression of numerical magnitude akin to that observed in behavior in the single-stream task. If anything, the results were suggestive of anti-compression ($k > 1$) in both tasks, although it should be noted that the improvement in model fit (relative to a linear model) in the single-stream task was small and not statistically significant (Fig. 3b, *left*).

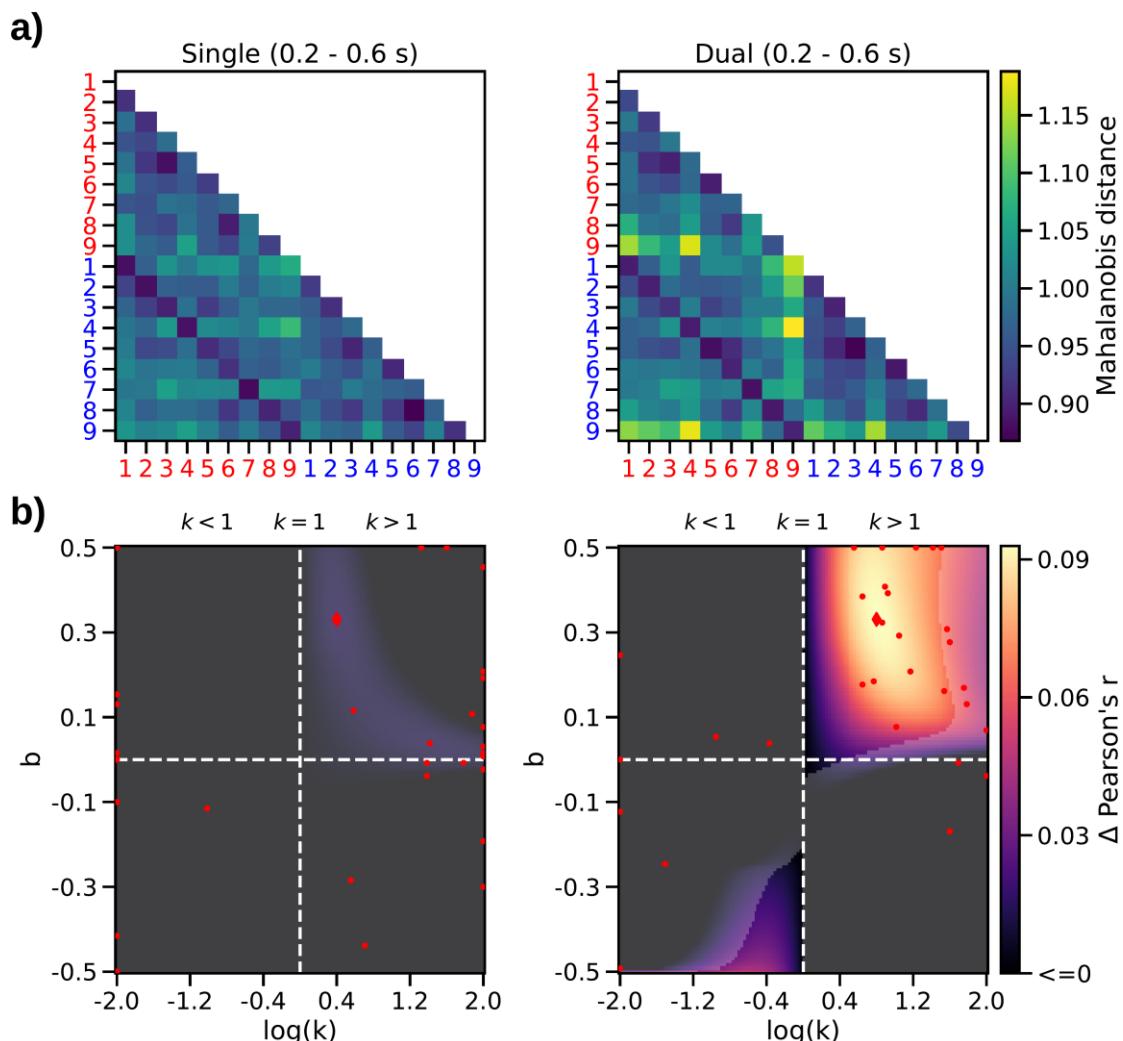


Figure 3. Neurometric RSA results. **a)** Mean ERP-RDMs in the time window of the numerical distance effect (see gray shading in Fig. 2b). **b)** Mean neurometric maps. *Left*: single-stream task; *right*: dual-stream task. Dashed lines indicate linear ($k = 1$) and unbiased ($b = 0$) parameterizations. Color scale indicates increase in model correlation (Δr) relative to the standard model where $k = 1$ and $b = 0$. Transparency

mask delineates where the increase was statistically significant ($p < 0.05$, FDR corrected). Red markers show maxima (diamond, mean map; dots, individual participant maps).

Univariate ERP results (CPP/P3)

We complemented our analysis by examining neurometric distortions also in univariate ERP signals, specifically in the sample-evoked CPP/P3 response. Previous research has implicated the CPP/P3 in decision formation, with its amplitude reflecting the perceived strength of evidence (Herding et al., 2019; O'Connell et al., 2012; Pisauro et al., 2017; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). CPP/P3 amplitudes were previously also found to be modulated by numerical sample values in the context of a dual-stream comparison task (Spitzer et al., 2017).

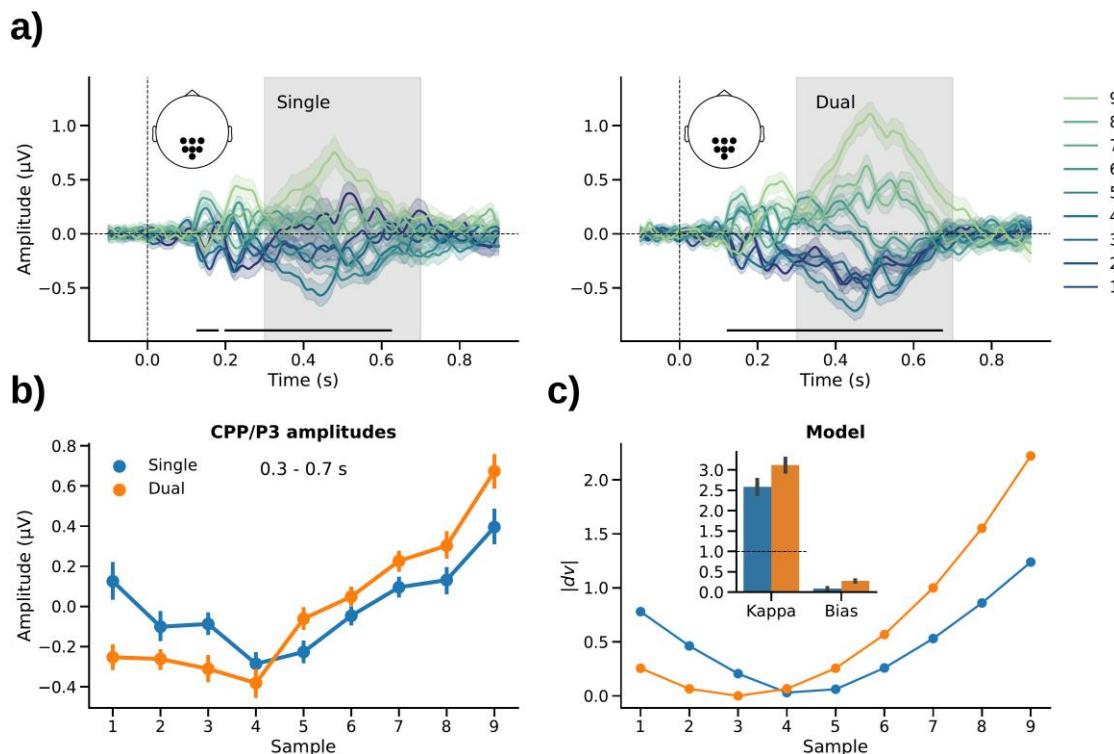


Figure 4. Univariate CPP/P3 results. a) Centro-parietal ERPs (mean-subtracted) evoked by sample values 1-9 in the single- (*left*) and dual-stream (*right*) tasks. Marker lines on bottom indicate time windows of significant differences between the 9 different sample values (all $p_{\text{cluster}} < 0.001$, repeated measures analysis of variance). b) Mean CPP/P3 amplitudes averaged over the time window outlined by gray shading in a (0.3 - 0.7 s). c) Neurometric model fit of CPP/P3 amplitudes. Line plot shows grand mean fit. Inset bar graph shows mean parameter estimates. Error indicators in all panels show SE.

We observed modulations of CPP/P3 amplitude by numerical value both in the single-stream (Fig. 4a, *left*; $p_{\text{cluster}} < 0.001$) and in the dual-stream task (Fig. 4a, *right*; $p_{\text{cluster}} < 0.001$, repeated measures analyses of variance), with the modulation in the dual-stream task appearing descriptively stronger. The mean amplitudes showed a U-shaped pattern over numbers 1-9 (Fig. 4b), consistent with previous findings that CPP/P3 reflects the strength of decisional evidence in an unsigned fashion, that is, a theoretical quantity similar to the *absolute* $|dv|$ in our psychometric

model, which would reflect the strength of evidence for either choice, “<” or “>” (see also Herding et al., 2019; O’Connell et al., 2012; Pisauro et al., 2017; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). Thus, for model-based analysis, we fitted Eq. 1 to the CPP/P3 amplitude data (using pairwise distance matrices and gridsearch analogous to our neurometric RSA above), but using $|dv|$ to generate the model-predicted pattern.

The best-fitting parameter estimates are shown in Fig. 4c (*inset bar graph*). Mirroring the RSA results, the parameter estimates based on CPP/P3 amplitude showed significant anti-compression ($k > 1$) in both tasks [single-stream: $M=2.58$, $SE=0.55$; $t(29)=2.88$, $p=0.007$, $d=0.53$, dual-stream: $M=3.12$, $SE=0.51$; $t(29)=4.15$, $p<0.001$, $d=0.76$, t-tests against 1], with no significant difference between tasks [$t(29)=0.69$, $p=0.49$, $d=0.18$, paired t-test]. A positive offset bias was again observed in the dual-stream task [$M=0.28$, $SE=0.06$; $t(29)=5.00$, $p<0.001$, $d=0.912$], but not in the single-stream task [$M=0.09$, $SE=0.05$; $t(29)=1.78$, $p=0.09$, $d=0.33$, t-tests against 0; difference: $t(29)=2.97$, $p=0.006$, $d=0.63$, paired t-test]. Together, the univariate ERP analysis thus corroborates our RSA finding that the neural processing of number samples in single-stream averaging was not characterized by compression but—if anything—by anti-compression, despite the evidence for compression in subsequent behavioral choice (cf. Fig. 1c-d).

Discussion

We observed opposite types of psychometric distortions (compression or anti-compression) in behavior when participants were instructed to process an identical stream of numbers in two distinct ways, namely, comparing the complete stream against a fixed target value or comparing two sub-streams against each other. However, contrary to expectations based on past research, the neural signals associated with the processing of the individual number samples showed evidence for anti-compression (i.e., enhanced processing of extreme values) under both instructions, regardless of whether extreme values were over- or underweighted in subsequent behavioral choice. We further observed qualitative differences between the sample representations in the two tasks, with a relatively weaker encoding of the samples’ numerical magnitude in the single-stream task.

In psychophysical research concerned with the how attributes of a physical stimulus (e.g., size, weight, color) relate to their subjective experience or perception, human observers are commonly found to underweight extreme values. Such subjective “compression” of magnitude can be observed in a great variety of settings, from basic sensory-perceptual judgments (Fechner, 1860; Stevens, 1957) to economic decisions (Kellen et al., 2016; McAllister & Tarbert, 1999; Tversky & Kahneman, 1992). There exist various theoretical accounts for the origin and potential benefits of subjective compression in perception and decision making (Bhui & Gershman, 2018; Ciranka et al., 2022; de Gardelle & Summerfield, 2011; Li et al., 2017; Pardo-Vazquez et al., 2019; Stewart et al., 2006; Summerfield & Li, 2018; Vandormael et al., 2017). Our present results seem to contrast these vast literatures with the finding that human brain signals tended to reflect the magnitude of numerical values in an *anti*-compressed fashion, even when they were compressed in later choice.

However, while the typical finding in many tasks is compression, there are task contexts where observers overweight extreme samples in choice behavior, in line with an *anti*-compression of sample values (Clarmann von Clarenau et al., 2022; Kunar et al., 2017; Ludvig et al., 2014; Luyckx et al., 2019; Shevlin et al., 2022; Spitzer et al., 2017; Tsetsos et al., 2012; Vanunu et al., 2020). We recently showed that such anti-compression can be beneficial in tasks that are computationally challenging (like our dual-stream comparison task), and where capacity-limited observers may be forced to selectively focus on a subset of the samples at the expense of others (Clarmann von Clarenau et al., 2022; see also Tsetsos et al., 2016). Our present findings may suggest that in task contexts in which higher-level processing capacities are exceeded, participant behavior might more directly reflect the brain's default response to extreme values (i.e., privileged processing).

Despite the lack of evidence for different neural geometries of numerical magnitude in the two tasks, the samples' overall representation in neural signals yet differed. In the dual-stream task, neural signals encoded the numerical magnitude of a sample more strongly than in the single-stream task. This result was unexpected because nominally, the numerical magnitude of a sample was of equal relevance in both tasks. In the single-stream task, in turn, we found relatively stronger encoding of which unique number symbol was presented. Further, although the color of a sample (red/blue) was task-irrelevant in the single-stream task, the neural encoding of color was as strong and as sustained as in the color-based dual-stream task. While these results were unexpected, they may suggest more general differences in the role of "abstract" magnitude processing (Dehaene, 2003; Nieder, 2005; Nieder & Dehaene, 2009; Piazza & Izard, 2009; Walsh, 2003) in the two tasks. Potentially, in the more challenging dual-stream task participants relied more directly on an intuitive "sense of magnitude" (Leibovich et al., 2017; Piazza et al., 2006; Spitzer et al., 2014) to gauge a sample's decision value. The computationally simpler single-stream task, in contrast, may have allowed them to engage in more symbolic-analytic processing (e.g., approximate arithmetics and/or verbalization) – processes that might be less amenable to EEG-decoding than the numerical distance pattern that prevailed in the dual-stream task.

The neural anti-compression of sample values in both of the tasks was also evident in univariate CPP/P3 signals which had previously been implicated in the decisional evaluation of stimulus information (Herding et al., 2019; O'Connell et al., 2012; Pisauro et al., 2017; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). Interpreting the amplitude of CPP/P3 signal as an index of the perceived strength of evidence, our findings in the single-stream task show a mismatch between the pattern observed in sample-by-sample processing (anti-compression, as revealed by neural data) and that in eventual judgment of the aggregate stream (compression, as evident in behavior). Future work will be required to identify the neural mechanisms underlying the eventual downweighting of extreme values in such task contexts, despite enhanced encoding in sample-level decision signals. It should be noted that the amplitude of P3 signals is known to be modulated by various factors, including how rare or surprising an event is (Donchin, 1981; Duncan- Johnson & Donchin, 1977). However, the sample values in our experiment were uniformly distributed, that is, each value occurred equally often on average, ruling out an explanation in terms of stimulus frequency.

Whether extreme values are over- or underweighted has major implications for the behavioral choices people make. One and the same numerical evidence, as in the present study, may lead to opposite choices, depending on how people respond to and process extreme values. For instance, anti-compression can explain systematic violations of “rational” axioms, such as transitivity, in multi-attribute choice (Summerfield & Tsetsos, 2015; Tsetsos et al., 2016). More generally, non-linear distortions of objective data (such as numbers) have often been interpreted as paradigmatic manifestations of seemingly “irrational” human behaviors in decision making (e.g., non-linear probability weighting in risky choice as assumed in cumulative prospect theory; Tversky & Kahneman, 1992). However, over the past years a new literature has evolved that recasts these behaviors as well-adapted policies of capacity-limited observers, fostering rather than hampering their measurable performance under these constraints (Bhui et al., 2021; Gigerenzer et al., 2011; Gigerenzer & Brighton, 2009; Juechems et al., 2021; Lieder & Griffiths, 2020; Sims, 2003, 2010; Tsetsos et al., 2016). Here, we shed new light on the open question of how such adaptive distortions may arise mechanistically, in terms of the neural signal patterns evoked by samples of evidence while observers are in the process of reaching a decision. Our finding of a “default” anti-compression of values in neural responses—regardless of subsequent behavior—raises the question at which exact processing stage value compression emerges, and how it leads to, for instance, the well-known “diminishing sensitivity” to larger values in economic choices (e.g., Tversky & Kahneman, 1992).

Materials and Methods

Participants

Thirty-two healthy volunteers took part in the experiment. We excluded two participants who reported having misunderstood the task instructions and who performed near chance level (50% correct choices) in one of the tasks (52% and 53%, respectively; both $p > 0.4$, Binomial tests against 0.5). Results are reported for the $n=30$ remaining participants (15 male, 15 female; mean age 27.4 ± 4.9 years; one left handed). All participants provided written informed consent and received €10 per hour as compensation, in addition to a €10 flat fee for participation, as well as a performance-dependent bonus ($€7.03 \pm 1.08$ on average). The study was approved by the ethics committee of the Max Planck Institute for Human Development.

Experimental design

Each participant performed two variants of a sequential number integration task (Clarmann von Clarenau et al., 2022; Spitzer et al., 2017). The stimulus protocols in the two task variants were identical (Fig. 1a). On each trial, participants viewed a sequence of 10 Arabic digits (randomly drawn from a uniform distribution of numbers 1 to 9) displayed in either red or blue font color (randomly assigned to each sample, with the restriction that each sequence contained 5 red and 5 blue samples). In the “averaging” (single-stream) task, participants were asked to judge whether the average of all 10 number samples in the sequence (regardless of their color) was larger or

smaller than 5. In the “comparison” (dual-stream) task, participants were asked to indicate whether the red or the blue samples had the higher average value. Past behavioral work has shown a psychometric compression of number values in the single-stream task, whereas anti-compression was evident in the latter dual-stream task (Clarmann von Clarenau et al., 2022; Spitzer et al., 2017).

The experiment was programmed in Python using the PsychoPy package (Peirce et al., 2019) and run on a Windows 10 PC. The experiment code is available on GitHub (https://github.com/sappelhoff/ecompr_experiment). Throughout the experiment, we additionally recorded eye-movements using an EyeLink 1000 Plus (SR Research Ltd., Canada), which were not analyzed in the present study. Participants were informed about the eye-movement recording and were instructed to keep their gaze at the center of the screen throughout the experiment.

Each trial started with a white central fixation stimulus (a combination of bulls eye and cross hair; Thaler et al. 2013) on an otherwise black screen. After 500 ms, the fixation stimulus disappeared and the number sequence was presented at a rate of 350 ms per sample (font Liberation Mono; height 3° visual angle; see Fig. 1a). Each sample was smoothly faded to black after 270 ms to improve the visual experience of the stimulus transitions. After the last sample, participants were prompted to enter a response by pressing the left or right button on a USB response pad (The Black Box ToolKit Ltd., UK). To avoid left/right motor response preparation during sequence presentation, in each of the two tasks, we randomized the mapping of responses (“smaller” or “larger” in the averaging task; “red” or “blue” in the comparison task) onto left/right button presses trial-by-trial, using a response screen (Fig. 1a, right).

If participants failed to respond within 3 s, the trial was discarded and after a delay of 100 ms, a message (“too slow!”) was displayed in red color for 1 s. On average, participants responded within 0.67 ± 0.27 s and timeouts occurred only on 0.03% of trials. On the remaining trials, performance feedback was displayed (“correct” or “wrong”, in green or orange color, respectively) for 350 ms. All feedback was displayed centrally in Liberation Mono font with a height of 1° visual angle. On 4.85% of trials, in which the objective sequence average was precisely 5 (in the single-stream task) or identical for red and green samples (in the dual-stream task), a random feedback message was displayed. These trials were excluded from the analysis of accuracy levels, but were included in the modeling- and EEG analyses (see also Spitzer et al., 2017). After feedback, the central fixation stimulus re-appeared and after 500 to 1500 ms (randomly varied), the next trial started.

Each participant first performed 300 trials in one of the tasks (single-stream averaging or dual-stream comparison), followed by 300 trials in the other task (in counterbalanced serial order across subjects). Thus, 3000 number samples were presented in each task condition and participant. Trials were performed in blocks of 50, with summary performance feedback (percentage correct choices) being provided after each block. After completing all blocks of the first task, participants received the instructions for the second task. To avoid differences in stimulus input, the second task was performed on the exact same number sequences as the first

task. Upon completing the second task, participants received a monetary bonus depending on their mean accuracy in both tasks.

EEG recording

The experiment was performed in an electrically shielded and soundproof cabin. Scalp EEG was recorded with 64 active electrodes (actiCap, Brain Products GmbH Munich, Germany) positioned according to the international 10% system. Electrode FCz was used as the recording reference. We additionally recorded the horizontal and vertical electrooculogram (EOG) and electrocardiogram (ECG) using passive electrode pairs with bipolar referencing. All electrodes were prepared to have an impedance of less than 10 kΩ. The data were recorded using a BrainAmp DC amplifier (Brain Products GmbH Munich, Germany) at a sampling rate of 1000 Hz, with an RC high-pass filter with a half-amplitude cutoff at 0.016 Hz (roll-off: 6 dB/octave) and low-pass filtered with an anti-aliasing filter of half-amplitude cutoff 450 Hz (roll-off: 24 dB/octave). The dataset is available on GIN in source format and formatted according to the Brain Imaging Data Structure (BIDS) using MNE-BIDS (Appelhoff et al., 2019; Gorgolewski et al., 2016; Pernet et al., 2019): https://gin.g-node.org/sappelhoff/mpib_ecomp_sourcedata/.

Behavioral data analysis

We calculated model-free decision weights to examine how strongly each numerical sample value (1, 2, ..., 9) contributed to participants' choices in the two tasks. In the single-stream task, these weights were computed as the proportion of times a sample value was associated with the subsequent choice "larger". Analogously, in the dual-stream task, the weights were computed as the proportion of times the sample's color (i.e., red or blue) was subsequently chosen (see also Spitzer et al., 2017). For comparison with model predictions (see below), we computed decision weights also from the model-predicted choice probabilities (CP , see Eq. 3) obtained from using the best fitting parameter estimates in each participant.

Psychometric model

To quantify psychometric distortions (i.e., compression or anti-compression) in behavior, we used a simple psychometric model that has been used extensively in previous work (Appelhoff et al., 2022; Clarmann von Clarenau et al., 2022; Li et al., 2017; Luyckx et al., 2019; Spitzer et al., 2017). The model formalizes the transformation of objective sample values X (here: numbers 1-9, normalized to the range [-1, 1]) into a subjective decision value dv as a sign-preserving power function:

$$dv = \frac{X+b}{|X+b|} \times |X+b|^k, \quad \text{Eq. 1}$$

where exponent k (kappa) determines the overall shape of the transformation ($k < 1$: compression; $k = 1$: linear; $k > 1$: anti-compression). Parameter b (bias) implements an overall weighting bias towards smaller ($b < 0$) or larger ($b > 0$) numbers. Sample-level decision values (dv) are integrated into a trial-level decision value (DV) by summation over samples:

$$DV = \sum_{i=1}^{10} d\nu_i \times c_i, \quad \text{Eq. 2}$$

where c is an indicator variable denoting a sample's color (red: -1, blue: +1) in the dual-stream task. In the single-stream task, c was fixed at 1. This way, Eq. 2 effectively implements a comparison between streams in the dual-stream task, and simple averaging in the single-stream task. Finally, the trial level decision value (DV) is transformed into a choice probability according to a logistic function:

$$CP = \frac{1}{1 + e^{\frac{-DV}{s}}}, \quad \text{Eq. 3}$$

where CP denotes the probability of choosing “>5” (in the single-stream task) or “blue>red” (in the dual-stream task), and parameter s quantifies the level of decision noise, with larger values of s implying more random choices.

The model was fitted to each participant's individual choice data using the Nelder-Mead method as implemented in SciPy (Virtanen et al., 2020), with parameter values restricted to the ranges (k : [0, 5]; b : [-0.5, 0.5], s : [0.01, 3]). Fitting was performed iteratively using 900 combinations of different starting values for each task condition, and the solution with the lowest Bayesian Information Criterion (BIC) was used in the analysis. Statistical analysis of the fitted parameters proceeded with conventional inferential tests on the group level.

EEG preprocessing

We used functions from MNE-Python (Gramfort et al., 2013) and PyPrep (Appelhoff et al., 2018; based on Bigdely-Shamlo et al., 2015) to automatically mark noisy segments and bad channels in the EEG recordings. We additionally screened all recordings visually to reject noisy segments or bad channels that the automatic procedures had missed. This way, on average, 2.5 ± 1.6 channels were discarded per participant. Next, we corrected ocular and cardiac artifacts using independent component analysis (ICA). To this end, we high-pass filtered a copy of the raw data at 1 Hz and downsampled it to 100 Hz. We then ran an extended infomax ICA on all EEG channels and time points that were not marked as bad in the prior inspection. Using the EOG and ECG recordings, we identified stereotypical eye blink, eye movement, and heartbeat artifact components through correlation with the independent component time courses. We visually inspected and rejected the artifact components before applying the ICA solution to the (Winkler et al., 2015). We then filtered the ICA-cleaned data between 0.1 and 40 Hz, interpolated bad channels, and re-referenced each channel to the average of all channels.

Event-related potentials (ERPs)

We epoched the preprocessed data from -0.1 to 0.9 s relative to each sample stimulus onset. Remaining bad epochs were rejected using a thresholding approach from the FASTER pipeline (Step 2; Nolan et al., 2010). On average, $n = 5764$ clean epochs (96.1%) per participant were retained for analysis. The epochs were then downsampled to 250 Hz and baseline corrected relative to the period from -0.1 to 0 s before stimulus onset. Since our analyses focused on stimulus-specific effects, we subtracted the overall mean waveform from the individual epochs, in

each of the two task conditions. The mean-subtracted epochs were then averaged into stimulus-specific ERPs for each sample value (1, 2, ..., 9) in each color (red/blue). Note that the individual samples in a stream were statistically independent by design, allowing us to examine stimulus-specific ERP responses in a time window that overlapped with the onset of the next sample stimulus.

Representational similarity analysis (RSA)

We used representational similarity analysis (RSA; Kriegeskorte & Kievit, 2013) to examine the encoding of sample information in multivariate ERP patterns. Specifically, we examined the representational geometry of our stimulus space (numbers 1 to 9, colored red or blue) in terms of the multivariate (dis-)similarity between the ERP topographies (64 channels) associated with the 18 different stimuli. Representational dissimilarity was computed as the Mahalanobis distance, between each pair of stimuli, yielding an 18x18 representational dissimilarity matrix (RDM), at each time point of the peri-sample epoch. To compute the Mahalanobis distance, we fitted a general linear model to the z-scored trial data, with each stimulus type specified as a condition, and used the residual trial-by-trial variance for pairwise distance calculation. This procedure ensured multivariate noise normalization for the RDMs (Guggenmos et al., 2018). Below, we refer to the thus obtained RDMs as ERP-RDMs.

To examine the information encoded in the ERP-RDM time courses, we used three different model RDMs (see Fig. 2a) reflecting (i) the unique digit symbols, with minimum dissimilarity between identical digits, and maximum dissimilarity between distinct digits (“digit” model), (ii) the samples’ color, with minimum (maximum) dissimilarity between same (different) colors (“color” model), and (iii) the numerical distance between samples, that is, the arithmetic difference between their objective number values (“numerical distance” model). To render the three models fully independent, we recursively orthogonalized each model RDM with respect to all others using the Gram-Schmidt process (Appelhoff et al., 2022; Spitzer et al., 2017). Finally, we assessed the match between each model RDM and the empirically observed ERP-RDMs via Pearson correlation at each time point, using only the lower triangle of the RDMs and omitting the diagonal, to exclude redundant matrix cells.

For statistical analysis of the RSA time courses, we used t-tests against zero with cluster-based permutation testing to control for multiple comparisons over time points (Maris & Oostenveld, 2007). To test whether RSA results differed between the single- and dual-stream tasks, we first computed their difference, followed by cluster-based permutation tests against zero. All permutation tests were performed over 1000 iterations with a cluster-defining threshold of $p = 0.01$ and cluster length as the critical statistic (thresholded at $p=0.01$).

Analysis of neurometric distortions

The theoretical model underlying conventional RSA of numerical distance effects (see above) is a straight number line, where the numbers (1, 2, ..., 9) are equidistant. The standard numerical distance model (Fig. 2a, *right*) is equivalent to a model of dv according to Eq. 1 (see *Psychometric*

model) where $k = 1$ and $b = 0$. To examine potential nonlinear distortions of the number representations in neural signals (“neurometric” distortions), we constructed numerical distance models based on $d\nu$ while varying the values of k (from $\log(k) = -2$ to $+2$, see Fig. 3b) and b (from -0.5 to 0.5). Varying k on a log scale centered around $\log(k) = 0$ (i.e., $k = 1$) ensured that parameter estimates were not biased to show anti-compression (or compression) by chance in subsequent gridsearch. For each parameter combination, we correlated the resulting model RDMs with the ERP-RDM, yielding a grid (“neurometric map”) of the parameter space (see Fig. 3b). The parameter combination with the maximum correlation was used as the estimate of the participant’s neurometric distortion parameters (see also Spitzer et al., 2017; Appelhoff et al., 2022). Statistical analysis of the neurometric parameter estimates proceeded with conventional statistical tests on the group level.

Univariate ERP analysis

For complementary inspection of univariate ERP responses evoked by the number samples, we examined the stimulus-specific ERP (see above) for each sample value (1-9; collapsed across red/blue colors). To focus on CPP/P3 responses (see *Results*), the ERPs were pooled over centro-parietal channels (CP1, P1, POz, Pz, CPz, CP2, P2) and amplitudes were examined in a time window from 300 ms to 700 ms based on previous work (Appelhoff et al., 2022; Polich, 2007; Spitzer et al., 2017; Wyart et al., 2015). The ERP time courses were analyzed statistically using cluster-based permutation testing (see above). For model-based analysis, we used the same approach as in our analyses of neurometric distortions in RSA (see above), except that the model RDMs were constructed from $|d\nu|$ (i.e., the absolute, unsigned magnitude of $d\nu$, see *Results*) and correlated with the pairwise differences in univariate ERP amplitude between samples 1-9.

Data availability

All data is available on GIN: https://gin.g-node.org/sappelhoff/mpib_ecomp_sourcedata/.

Code availability

All analysis code is available on GitHub: https://github.com/sappelhoff/ecomp_analysis. The experiment code is available on GitHub: https://github.com/sappelhoff/ecomp_experiment.

Ethics information

The study was approved by the ethics committee of the Max Planck Institute for Human Development.

Acknowledgements

We thank Anna Faschinger, Gabriele Inciuraite, Aleksandra Zinoveva, Larissa Samaan, Simon Ciranka, and Jann Wäscher for help with data collection, and Verena Clarmann von Clarenau and Thorsten Pachur for helpful discussions.

Author contributions

SA, BS: Formal analysis, Visualization, Methodology, Conceptualization, Project Administration, Writing - original draft

SA: Investigation, Validation, Data curation, Software

SA, BS, RH: Writing - review & editing

BS: Supervision

RH, BS: Resources

Funding

This work was supported by a European Research Council Consolidator Grant ERC-2020-COG-101000972 (BS).

Competing interests

The authors declare no competing interests.

References

Appelhoff, S., Hertwig, R., & Spitzer, B. (2022). Control over sampling boosts numerical evidence processing in human decisions from experience. *Cerebral Cortex*.
<https://doi.org/10.1093/cercor/bhac062>

Appelhoff, S., Hurst, A. J., Lawrence, A., Li, A., Mantilla Ramos, Y. J., O'Reilly, C., Liang, X., & Dancker, J. (2018). *PyPREP: A Python implementation of the preprocessing pipeline (PREP) for EEG data*. <https://doi.org/10.5281/zenodo.6363575>

Appelhoff, S., Sanderson, M., Brooks, T. L., Vliet, M. van, Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A. P., Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, 4(44), 1896. <https://doi.org/10.21105/joss.01896>

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22(1), 23. <https://doi.org/10.2307/1909829>

Bhui, R., & Gershman, S. J. (2018). Decision by sampling implements efficient coding of psychoeconomic functions. *Psychological Review*, 125(6), 985–1001.
<https://doi.org/10.1037/rev0000123>

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, 41, 15–21. <https://doi.org/10.1016/j.cobeha.2021.02.015>

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00016>

Ciranka, S., Linde-Domingo, J., Padezhki, I., Wicherz, C., Wu, C. M., & Spitzer, B. (2022). Asymmetric reinforcement learning facilitates human inference of transitive relations. *Nature Human Behaviour*, 1–10. <https://doi.org/10.1038/s41562-021-01263-w>

Clarmann von Clarenau, V., Pachur, T., & Spitzer, B. (2022). *Over- and Underweighting of Extreme Values in Decisions from Sequential Samples* [Preprint]. PsyArXiv.

<https://doi.org/10.31234/osf.io/6yj4r>

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment.

Proceedings of the National Academy of Sciences, 108(32), 13341–13346.

<https://doi.org/10.1073/pnas.1104517108>

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4), 145–147. [https://doi.org/10.1016/s1364-6613\(03\)00055-x](https://doi.org/10.1016/s1364-6613(03)00055-x)

Donchin, E. (1981). Surprise!? Surprise? *Psychophysiology*, 18(5), 493–513.

<https://doi.org/10.1111/j.1469-8986.1981.tb01815.x>

Duncan- Johnson, C. C., & Donchin, E. (1977). On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology*, 14(5), 456–467. <https://doi.org/10.1111/j.1469-8986.1977.tb01312.x>

Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel.

<https://books.google.de/books?id=6rINAAAAYAAJ>

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1(1), 107–143. <https://doi.org/10.1111/j.1756-8765.2008.01006.x>

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics*. Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780199744282.001.0001>

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044.

<https://doi.org/10.1038/sdata.2016.44>

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267).

<https://doi.org/10.3389/fnins.2013.00267>

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, 173, 434–447.

<https://doi.org/10.1016/j.neuroimage.2018.02.044>

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *NeuroImage*, 201, 116011. <https://doi.org/10.1016/j.neuroimage.2019.116011>

Juechems, K., Balaguer, J., Spitzer, B., & Summerfield, C. (2021). Optimal utility and probability functions for agents with finite computational precision. *Proceedings of the National Academy of Sciences*, 118(2). <https://doi.org/10.1073/pnas.2002232118>

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition*, 157, 126–138.

<https://doi.org/10.1016/j.cognition.2016.08.020>

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412.

<https://doi.org/10.1016/j.tics.2013.06.007>

Kunar, M. A., Watson, D. G., Tsetsos, K., & Chater, N. (2017). The influence of attention on value integration. *Attention, Perception, & Psychophysics*, 79(6), 1615–1627.

<https://doi.org/10.3758/s13414-017-1340-7>

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. <https://doi.org/10.1017/S0140525X16000960>

Li, V., Castañón, S. H., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLOS Computational Biology*, 13(8), e1005723. <https://doi.org/10.1371/journal.pcbi.1005723>

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43. <https://doi.org/10.1017/S0140525X1900061X>

Longo, M. R., & Lourenco, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia*, 45(7), 1400–1407. <https://doi.org/10.1016/j.neuropsychologia.2006.11.002>

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme Outcomes Sway Risky Decisions from Experience: Risky Decisions and Extreme Outcomes. *Journal of Behavioral Decision Making*, 27(2), 146–156. <https://doi.org/10.1002/bdm.1792>

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning. *eLife*, 8, e42816. <https://doi.org/10.7554/eLife.42816>

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>

McAllister, P., & Tarbert, H. (1999). Bargaining, utility and rents: Analysing the effect of potential lease termination on rent negotiation outcomes. *Journal of Property Investment & Finance*, 17(4), 353–364. <https://doi.org/10.1108/14635789910271746>

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, 6(3), 177–190. <https://doi.org/10.1038/nrn1626>

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, 32(1), 185–208. <https://doi.org/10.1146/annurev.neuro.051508.135550>

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1), 149–157.

[https://doi.org/10.1016/s0896-6273\(02\)01144-3](https://doi.org/10.1016/s0896-6273(02)01144-3)

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, 192(1), 152–162.

<https://doi.org/10.1016/j.jneumeth.2010.07.015>

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, 15(12), 1729–1735. <https://doi.org/10.1038/nn.3248>

Pardo-Vazquez, J. L., Castiñeiras-de Saa, J. R., Valente, M., Damião, I., Costa, T., Vicente, M. I., Mendonça, A. G., Mainen, Z. F., & Renart, A. (2019). The mechanistic foundation of Weber's law. *Nature Neuroscience*, 22(9), 1493–1502. <https://doi.org/10.1038/s41593-019-0439-7>

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, 6(1), 103. <https://doi.org/10.1038/s41597-019-0104-8>

Piazza, M., & Izard, V. (2009). How Humans Count: Numerosity and the Parietal Cortex. *The Neuroscientist*, 15(3), 261–273. <https://doi.org/10.1177/1073858409333073>

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, 1106(1), 177–188. <https://doi.org/10.1016/j.brainres.2006.05.104>

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, 8(1), 15808. <https://doi.org/10.1038/ncomms15808>

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>

Sheahan, H., Luyckx, F., Nelli, S., Teupe, C., & Summerfield, C. (2021). Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron*, 109, 1214–1226. <https://doi.org/10.1016/j.neuron.2021.02.004>

Shevlin, B. R. K., Smith, S. M., Hausfeld, J., & Krajbich, I. (2022). High-value decisions are fast and accurate, inconsistent with diminishing value sensitivity. *Proceedings of the National Academy of Sciences*, 119(6), e2101508119. <https://doi.org/10.1073/pnas.2101508119>

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3), 665–690. [https://doi.org/10.1016/S0304-3932\(03\)00029-1](https://doi.org/10.1016/S0304-3932(03)00029-1)

Sims, C. A. (2010). Chapter 4—Rational Inattention and Monetary Economics. In B. M. Friedman & M. Woodford (Eds.), *Handbook of Monetary Economics* (Vol. 3, pp. 155–181). Elsevier. <https://doi.org/10.1016/B978-0-444-53238-1.00004-1>

Spitzer, B., Blankenburg, F., & Summerfield, C. (2016). Rhythmic gain control during supramodal integration of approximate number. *NeuroImage*, 129, 470–479. <https://doi.org/10.1016/j.neuroimage.2015.12.024>

Spitzer, B., Fleck, S., & Blankenburg, F. (2014). Parametric Alpha- and Beta-Band Signatures of Supramodal Numerosity Information in Human Working Memory. *Journal of Neuroscience*, 34(12), 4293–4302. <https://doi.org/10.1523/JNEUROSCI.4580-13.2014>

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, 1(8), 1–8. <https://doi.org/10.1038/s41562-017-0145>

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64(3), 153–181. <https://doi.org/10.1037/h0046162>

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53(1), 1–26. <https://doi.org/10.1016/j.cogpsych.2005.10.003>

Summerfield, C., & Li, V. (2018). Perceptual suboptimality: Bug or feature? *Behavioral and Brain Sciences*, 41. <https://doi.org/10.1017/S0140525X18001437>

Summerfield, C., & Parpart, P. (2022). Normative Principles for Decision-Making in Natural Environments. *Annual Review of Psychology*, 73(1), 53–77.
<https://doi.org/10.1146/annurev-psych-020821-104057>

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *Trends in Cognitive Sciences*, 19(1), 27–34. <https://doi.org/10.1016/j.tics.2014.11.005>

Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with Magnetoencephalography: Evidence for a Shared Representation of Magnitude. *Journal of Cognitive Neuroscience*, 30(7), 999–1010.
https://doi.org/10.1162/jocn_a_01257

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, 109(24), 9659–9664. <https://doi.org/10.1073/pnas.1119569109>

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, 113(11), 3102–3107.
<https://doi.org/10.1073/pnas.1519157113>

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
<https://doi.org/10.1007/BF00122574>

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, 42(1), 1636–1643. <https://doi.org/10.1111/ejn.12936>

Vandormael, H., Herce Castañón, S., Balaguer, J., Li, V., & Summerfield, C. (2017). Robust sampling of decision information during perceptual choice. *Proceedings of the National*

Academy of Sciences, 114(10), 2771–2776. <https://doi.org/10.1073/pnas.1613950114>

Vanunu, Y., Hotaling, J. M., & Newell, B. R. (2020). Elucidating the differential impact of extreme-outcomes in perceptual and preferential choice. *Cognitive Psychology*, 119, 101274. <https://doi.org/10.1016/j.cogpsych.2020.101274>

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488. <https://doi.org/10.1016/j.tics.2003.09.002>

Winkler, I., Debener, S., Müller, K., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4101–4105. <https://doi.org/10.1109/EMBC.2015.7319296>

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. *Neuron*, 76(4), 847–858. <https://doi.org/10.1016/j.neuron.2012.09.015>

Wyart, V., Myers, N. E., & Summerfield, C. (2015). Neural Mechanisms of Human Perceptual Choice Under Focused and Divided Attention. *Journal of Neuroscience*, 35(8), 3485–3498. <https://doi.org/10.1523/JNEUROSCI.3276-14.2015>