PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Zénon A, Salvaggio S, Andres M. 2026 Pupil size variations reveal Bayesian inference in cognitive arithmetic. *Proc. R. Soc. B* **293**: 20251937.
https://doi.org/10.1098/rspb.2025.1937

Received: 29 July 2025
Accepted: 30 January 2026

Subject Category:
Neuroscience and cognition

Subject Areas:
behaviour, neuroscience, computational biology

Keywords:
numerical cognition, information theory, pupillometry, arousal, computational neuroscience

Author for correspondence:
Alexandre Zénon
e-mail: alexandre.zenon@u-bordeaux.fr

Pupil size variations reveal Bayesian inference in cognitive arithmetic

Alexandre Zénon[1], Samuel Salvaggio[2] and Michael Andres[2]

[1]CNRS, INCIA, UMR 5287, Université de Bordeaux, Bordeaux, Nouvelle-Aquitaine, France
[2]Psychological Sciences Research Institute, UCLouvain, Louvain-la-Neuve, Wallonia, Belgium

AZ, 0000-0001-7989-1261; MA, 0000-0002-7505-8157

The assumption that the brain relies on Bayesian inference has been successful in accounting for many behavioural and neurophysiological observations, but the dependence on such a mechanism in arithmetic remains unknown. Bayesian inference implies the representation of uncertainty and reliance on prior beliefs. In arithmetic problem-solving, it would consist of refining priors over the response range as the system progressively integrates information conveyed by the operands. To test this hypothesis, we designed three experiments in which participants computed the sum of two numbers presented sequentially through headphones. The first operand was either highly informative and contributed to narrowing down the response range or poorly informative and conveyed little information about plausible responses. Throughout all experiments, pupil-related arousal signalled the information gain associated with the first operand, indicating that participants updated the probability distribution of responses upon hearing that first stimulus. Moreover, when participants received more informative operands and when their pupils dilated more during this presentation, they were also faster to respond, indicating that greater initial information processing facilitated quicker problem-solving. These findings show that Bayesian inference is central to arithmetic problem-solving and that information gains consequent to the integration of the operands can be tracked over time through pupillometry.

1. Introduction

Bayesian inference is viewed as a core mechanism of the human brain, providing a comprehensive account of behavioural and neurophysiological observations [1]. Reliance on Bayesian inference implies that individuals tackle problems with some *a priori* knowledge (or probability distribution) of the expected responses and update these beliefs upon reception of new pieces of information [2]. Bayesian inference has proved effective in accounting for a wide range of human skills, from perception [3] to sensorimotor learning [4].

A challenge for these probabilistic theories is to account for the computations underlying culturally inherited functions, such as language or mathematics, which require the manipulation of abstract symbols according to formal rules [5]. In the present work, we propose to test the hypothesis that, besides symbolic manipulation by means of syntactic rules, Bayesian inference provides a complementary computational principle describing how the brain may dynamically integrate information and refine expectations during calculation. According to this view, formal symbolic operations would be embedded within a probabilistic system that continuously updates beliefs about possible outcomes, a process that could be quantitatively characterized through measures of information gain (IG). Just as priors guide the interpretation of ambiguous sensory input, they may guide mental arithmetic by narrowing down the range of likely outcomes based on operand values and past experience. Such reliance on prior distributions has been proposed to explain performance in addition [6], multiplication [7] as well as number comparison [8]. These models typically refer to statistical properties of arithmetic learning, such as frequency or order of presentation in textbooks, to account for the strength of problem representations in long-term memory [9,10]. They share the assumption that arithmetic problems encountered more often are processed more efficiently [11]. Such a processing advantage is compatible with the view that, through repeated practice, priors over responses have been adapted to fit frequent problems particularly well. But while frequency-based models posit that more practised problems are retrieved faster, Bayesian inference entails an active updating process. This distinction is critical: on top of explaining why more frequent problems are easier, Bayesian models explain how difficulty is dynamically reduced through the progressive narrowing of possible responses, thereby reducing uncertainty during problem-solving. Moreover, while probabilistic models are concerned with overlearned problems represented in an organized network in memory (e.g. $2 \times 4 = 8$), Bayesian inference can also account for the solving of more complex arithmetic problems (e.g. $43 + 8 = 51$) that involve greater structural variability and reduced familiarity, requiring procedural computations.

We aim to demonstrate the inherently probabilistic nature of arithmetic problem-solving by providing a mechanistic account of the relationship between arithmetic difficulty and IG, a measure of the divergence between prior and posterior beliefs that has recently been proposed as a universal currency to express cognitive cost [12]. This framework provides a theoretical basis for addressing the determinants of arithmetic difficulty and a quantitative assessment method leveraging the pupil response [13].

To address the question of the reliance on prior beliefs for arithmetic problem-solving, measuring response times (RTs) may not be sufficient, as they index the final outcome of all serial or parallel processes involved in the task, reflecting the sum of all computations from the encoding of the stimuli to the solving of the arithmetic problem. However, updating of evidence occurs in multiple, successive steps: encoding the first operand (e.g. 47), the operator (e.g. + or −) and the second operand (e.g. 4) allows refining progressively the prior distribution of the probabilities assigned to different candidate responses within a plausible range. In other words, each operand contributes to reducing the uncertainty about the response and is thus associated with a specific and quantifiable IG. In order to assess the independent contribution of each operand to the cognitive cost of solving arithmetic problems, we chose to measure changes in pupil dilation from the presentation of the first operand, thus long before a final decision could be made on the response. Besides light, pupil size is influenced by several affective and cognitive factors such as emotions, surprise or mental effort [14–16] and was also found to increase consecutively to the solving of arithmetic problems [17]. Neurophysiological models have linked these pupil signals to neuromodulatory systems (notably the locus coeruleus–noradrenaline system) and to cortical control structures implicated in meta-learning and belief regulation (e.g. dorsal anterior cingulate cortex (dACC)) [18,19]. Recent computational and empirical work has made a close link between pupil dynamics and model-based belief updating. For example, studies have shown that pupil-linked arousal co-varies with risk prediction error [20], inferred environmental volatility [21] or trial-by-trial adjustments of learning rates [22]. Computational models have proposed that dACC–brainstem ensembles coordinate belief updating and neuromodulatory control, mechanisms that can manifest in pupil dynamics [23]. These diverse accounts can be understood within a common quantitative framework based on IG, as formalized by Zénon [13]. In that model, the pupil signal is interpreted as reflecting the magnitude of belief updating, the Kullback–Leibler divergence between prior and posterior probability distributions, providing a unified metric for quantifying the cognitive cost associated with surprise, uncertainty reduction or volatility-driven adjustments in belief. We adopt this model here to test whether arithmetic problem-solving, such as perceptual and reinforcement learning, relies on Bayesian belief updating, whose informational cost can be directly tracked through pupillometry.

We conducted a series of experiments where the pupil size of adult participants was recorded while they solved addition problems with two-digit numbers presented sequentially in auditory format. The participants were informed about the range of all possible responses before the experiment, so that they all had the same prior internal models about the probable distribution of the responses. The information conveyed by the first operand was manipulated differently across conditions so that it could be more or less informative, thus more or less efficient in narrowing down the range of possible responses, to test the prediction that mental arithmetic works as a form of Bayesian inference (figure 1). Importantly, high IG trials entailed fewer plausible outcomes and thus placed lower demands on working memory (WM), allowing us to dissociate the effects of IG from memory load.

2. Experiment 1

The first experiment used addition problems whose operands consisted of a two-digit number and a single digit. In such problems, the two-digit number was highly informative because it greatly reduced the range of responses, compared with the single digit, which was less informative in that respect. The experiment was based on a within-subject design, with two sessions consisting of the same addition problems, except that the order of the operands was reversed between sessions to manipulate the IG associated with the first operand. In one session, the first operand was the two-digit number and the second operand was the single digit (xx + x). In the other session, the first operand was the single digit and the second was the two-digit number (x + xx). We predicted that, after hearing the first operand, pupil dilation should vary as a function of the IG. Specifically, given the format of the addition problems tested, pupil size should dilate more when the first operand is a highly informative two-digit number than when it is a single digit with low-informative value.

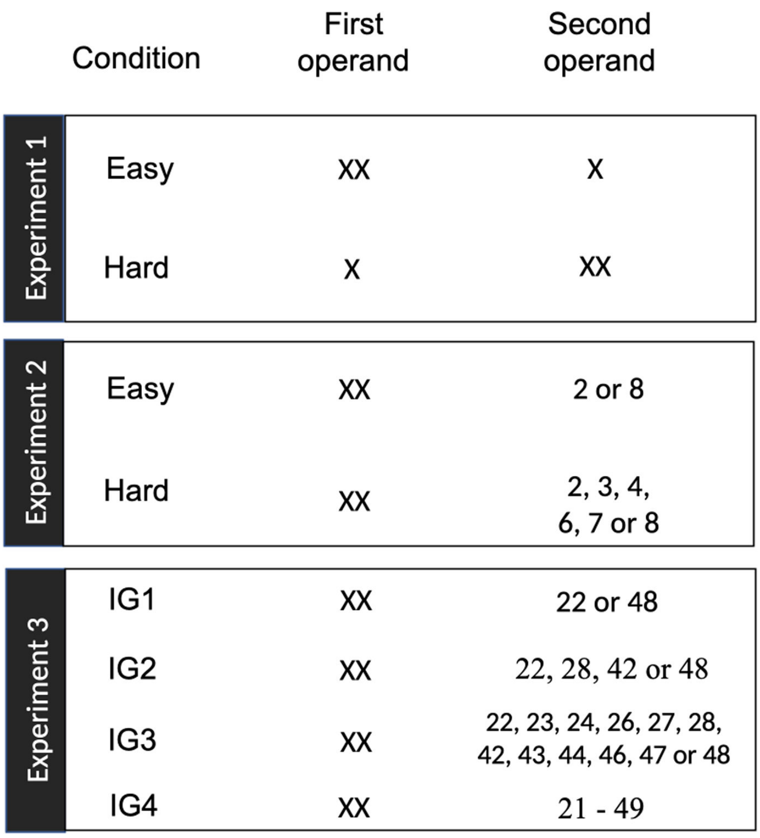

Figure 1. Schematic of the three experiments. The task consisted of a mental addition task, with the two operands being presented successively and auditorily. The difference across experiments was the choice of values for those operands.

(a) Method

(i) Participants

Twenty French-speaking participants took part in this experiment (nine males; two left-handed; mean age and standard deviation, M = 20.7, s.d. = 1.8). They all had normal or corrected-by-lenses vision and did not report any antecedent of mathematical learning disability when asked by the experimenter. The experiment was in accordance with the ethical standards established by the Declaration of Helsinki and was given prior approval by the local ethics committee. Owing to the lack of literature on the effect we sought to test, the sample size was arbitrarily pre-planned to have a balanced and sufficient number of participants for the various experimental conditions and randomizations.

The study complies with American Psychological Association ethical standards in the treatment of the acquired data.

(ii) Apparatus

The experiment was run on a PC equipped with a 22 inch LCD screen. The MATLAB software (v. 2018a) controlled the stimulus presentation and the recording of the verbal response (MathWorks Inc., 2018). An Eyelink 1000 tower-mounted camera was used to track eye movements (SR Research, Mississauga, Canada; sampling rate: 1000 Hz; average accuracy range: 0.25° angle to 0.5° angle; gaze tracking range of 32° angle horizontally and 25° angle vertically). Participants were placed at 60 cm from the screen. Prior to each experimental block, the eye tracker was calibrated to the screen using a built-in nine point protocol.

(iii) Stimulus materials

The auditory stimuli consisted of either double-digit French number words whose magnitude ranged from 21 to 69, excluding numbers ending in '0' (i.e. 30, 40, 50, 60) or single-digit numbers from '1' to '9'. The number words were recorded in a stereo audio file whose duration was adjusted to 750 ms for two-digit numbers and 500 ms for one-digit numbers. The screen stayed unchanged for the duration of all the experiments. It consisted of a central white fixation dot on a grey background. The experiment counted 144 trials for each condition, giving 288 trials in four separate blocks. Thus, there were two blocks where the stimulus order consisted of a double digit followed by one digit, and two blocks where the stimulus order was a single digit followed by a double digit. We used the Kullback–Leibler (KL) divergence between prior and posterior beliefs to quantify the IG associated with the first operand, both in the double-digit first blocks and in the single-digit first blocks. The KL divergence can be calculated, for probability distribution P and Q, as shown in equation (3.1).

$$\mathrm{KL}(\mathrm{P}\,|\,\mathrm{Q}) = \sum P(x)\log\frac{P(x)}{Q(x)}. \tag{3.1}$$

In the present experiment, the KL divergence of the two-digit-number-first (high-informative) condition was 2.38 bits, while the KL divergence of the one-digit-number-first (low-informative) condition was 0.24 bits.

(iv) Task and procedure

Following a 1000 ms fixation period, each trial started with a number word that was played in the headphones while the screen stayed unchanged. Then, 2750 ms after trial onset, the second number was played in the headphones. Participants had to add the two numbers and answer orally. The display remained in this configuration for an additional 3750 ms while participants' pupil diameter continued to be recorded. The next trial began immediately afterward. Participants were instructed to constantly look at a central fixation cross, which was continuously displayed on screen throughout each trial, including during auditory stimulus presentation. Before starting all blocks of trials, they were reminded of the range of the possible response and the order of stimuli (i.e. double-digit or single-digit first). The order of the blocks was counterbalanced across all participants.

(v) Data analysis

Erroneous trials were excluded from analyses (10.3% of the dataset; no difference across conditions: Friedman test, $\chi^2 = 2.88$, $p = 0.09$). RTs corresponded to the delay between the onset of the second operand and the onset of the subject's response. Pupillary data were pre-processed by linearly interpolating missing data resulting from eye blinks and by high-pass filtering it (digital Chebyshev filter, 1/16 Hz cut-off frequency). Time-series analyses of pupil dilation aligned to first operand onset (see electronic supplementary material, figures S1–S3) confirmed that effects emerged within the typical 0.5–1.5 s latency window, suggesting that pupil changes preceded response preparation and reflected initial evidence accumulation. We then isolated the impulse response of the pupil to cognitive events (operand presentations and responses) by running an autoregressive model with exogenous inputs (ARX) [24]. The orders of the ARX model were fixed to 3 (autoregressive process) and 15 (exogenous inputs), whereas the delay parameter was chosen within a range of 1–5, so as to maximize the peak of the impulse response. The pupillary response to the operand presentations and to the response in each trial was then obtained by running a generalized linear model (GLM) with the pupil signal as the explained variable and three regressors per trial, each obtained from the convolution of the event vectors (zero everywhere except at the time of occurrence of the events) with the subject-specific pupillary impulse response. This resulted in one regression coefficient per event per trial, which represented the magnitude of the response to that event in that trial. The autocorrelation in the residuals was taken care of with a second-order autoregressive model model [25]. We then ran generalized linear mixed models (GLMMs) with the coefficients corresponding to the first operand as dependent variable and IG condition (2.38 bits versus 0.24 bits) as independent variable: Pupil size ~ 1 + IG + (1 + IG | Participant). The dependent variable was transformed so as to make its distribution closer to normal (see equation (3.2)).

$$x = \sqrt{x} \text{ if } x > 0; \quad x = -\sqrt{-x} \text{ if } x < 0. \tag{3.2}$$

Residuals were systematically inspected by means of quantile-quantile plots and per-participant boxplots. Degrees of freedom were estimated with Satterthwaite's approach, and estimation of model parameters was performed with the restricted maximum likelihood (REML) method. We also ran a GLMM on log-transformed RTs with IG condition and z-scored pupil dilation as dependent variables: RTs ~ 1 + pupil + IG + (1 + pupil + IG | Participant).

All analyses were conducted with MATLAB 2018a (MathWorks Inc., 2018) and Jamovi (The Jamovi project (2021), Jamovi v. 1.6 (computer software), retrieved from https://www.jamovi.org).

(vi) Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations and all measures in the study. All data and statistical models are available at https://researchbox.org/3762. Data were analysed using MATLAB v. R2016b (MathWorks Inc.; code available at https://github.com/alexandre-zenon) and Jamovi v. 2.5.4.0 (The Jamovi project).

(b) Results

We found that the IG condition affected pupil dilation significantly, with larger pupil responses being associated with larger IG ($F_{1,19.2} = 14.3$, $p = 0.001$, β mean (95% CI): 0.155 (0.0744–0.2348); see figure 2A), in agreement with our hypothesis. The RTs were found to depend also on condition ($F_{1,19.1} = 30.6$, $p < 0.001$, β mean (95% CI): −0.2452 (−0.332 to −0.1583)), with longer RTs in trials with large IG, though these conditions corresponded to trials with double-digit operands, whose presentation lasted 250 ms longer. This makes it difficult to determine whether the effect of the condition on RTs was owing to information content or to the duration of the presentation. However, and more importantly, we found that larger pupil dilation to the first operand led to shorter RTs ($F_{1,5279.7} = 76.5$, $p < 0.001$, β mean (95% CI): −0.2654 (−0.325 to −0.206); see figure 2A), confirming that the amount of information being processed during first operand presentation was instrumental in decreasing computation time at the moment of second operand presentation.

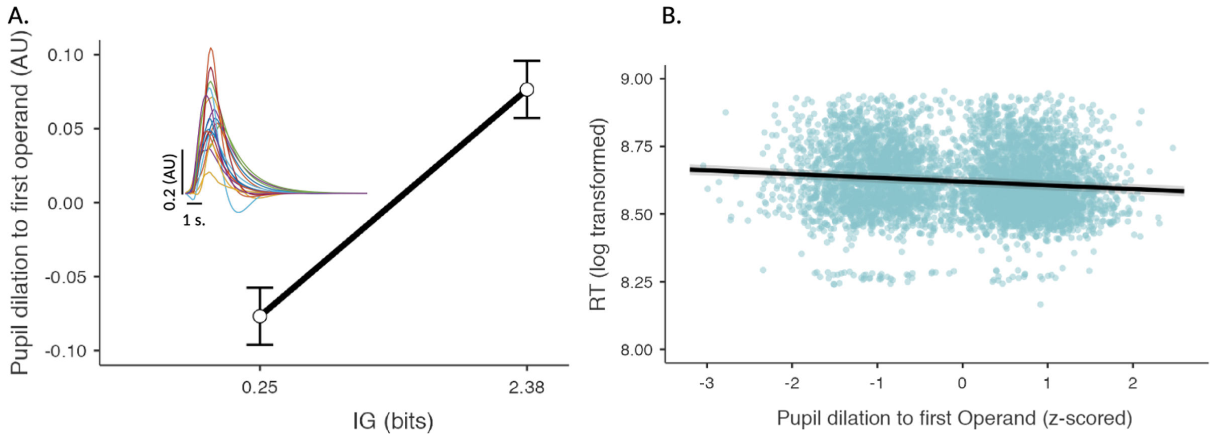

Figure 2. (A) Average pupil dilation to first operand presentation as a function of its information gain in bits. Error bars indicate the standard error of the mean. Inset in the upper left corner shows individual impulse responses to operand onset. (B) RT (log-transformed) as a function of pupil dilation to the first operand. The confidence interval of the slope is indicated in the grey shade. Each dot corresponds to an individual trial.

(c) Interim discussion

In this experiment, participants' pupil size was recorded while they were solving addition problems. Each operand was presented sequentially, and the amount of information provided by the first operand was directly manipulated. Results showed that the pupil response to the presentation of the first operand was significantly higher in the 2.38 bits condition when compared with the 0.24 bits condition. The analysis of RTs showed that problems where the first operand was less informative were also more difficult and took longer to be resolved. We also highlighted a negative relationship between average pupil size and RT. These results converge to show that pupils dilated in relation to the IG associated with the first operand, which helped participants to narrow down the range of possible solutions beforehand, helping them to solve the arithmetic problem faster. They are consistent with the hypothesis that arithmetic problem-solving is based on probabilistic inference and relies on progressive updating of prior distributions over possible responses, as new information is received.

Past research has long been animated by the question of which problems are solved by procedural strategies and which are resolved through memory retrieval strategies [11,26,27]. Procedural strategies, such as counting or decomposition, were initially associated with longer response latencies and higher error rates, compared with retrieval strategies, underlining the role of arithmetic difficulty in the selection of the solving strategy [28]. However, previous studies mainly investigated the solving of single-digit arithmetic problems [29]. In comparison, little is known about the difficulty of two-digit arithmetic problem-solving. Conventionally, difficulty indicators consist of categorical criteria, such as the total number of digits or the need to carry over digits across the tens places, which have been used to account for increased RTs and error rates [30–33]. However, these indicators are limited since they fail to provide an operational definition of arithmetic difficulty for all problems—only for categories—and by extension, they fail to provide a comprehensive account of the cognitive cost of arithmetic problem-solving at a more conceptual level. The present results provide a complementary account of two-digit arithmetic difficulty, based on the progressive narrowing down of the range of solutions, quantified through IG.

While these results support our interpretation, an alternative explanation could be that pupil dilation reflected differences in phonological or lexical processing, or WM load owing to the operand's length. Indeed, the encoding of the first operands differed in phonological and lexical complexity between the two lists, and the auditory presentation of a two-digit number was 250 ms longer than the auditory presentation of a single digit, resulting in a higher WM load in trials involving two-digit numbers. We therefore conducted a second experiment where the first operands were the same, but their informative value changed as a function of the second operands used in each list of problems, with the aim of replicating the effect of the information conveyed by the first operand on pupil dilation while excluding other sources of arousal increase.

3. Experiment 2

To counteract possible biases in pupil size measures owing to phonological or lexical differences related to the hearing of single-digit versus two-digit numbers as first operands, we conducted a second experiment where the first operand was always a two-digit number taken from the same list in both the low and the high IG conditions. The informative value of the first operand in these two conditions was actually given by the relative value of the second operand: in the high IG condition, the second operand was chosen in a list of two possible digits, whereas in the low IG condition, it was chosen in a list of six possible digits, leaving more uncertainty about the arithmetic outcome. We predicted a larger pupil dilation after hearing the first operand in the high IG condition, where the second operand carried little additional information about the outcome of the calculation.

(a) Method

(i) Participants

Twenty French-speaking participants took part in this experiment (eight males; one left-handed; $M = 22.2$, s.d. = 3.2). They all had normal or corrected-by-lenses vision and did not report any antecedent of mathematical learning disability when asked by the experimenter. The experiment was in accordance with the ethical standards established by the Declaration of Helsinki and was pre-approved by the local ethics committee.

(ii) Stimulus materials

The first operands consisted of double-digit French number words whose magnitude ranged from 21 to 49, excluding numbers ending in '0' (i.e. 20, 30, 40). The second operands consisted of either {'2', '8'} in the high IG condition or {'2', '3', '4', '6', '7', '8'} in the low IG condition. The experiment counted 162 trials for each condition and location of the bright/dark part of the screen, giving 324 trials in six separated blocks. In the present experiment, the KL divergence of the high IG condition was 2.81 bits, while the KL divergence of the low IG condition was 1.7 bits.

(iii) Task and procedure

The procedure was the same as in experiment 1. Before starting all blocks of trials, participants were reminded of the range of the possible responses and of the second operand. The order of the blocks was counterbalanced across all participants.

(iv) Data analysis

Two subjects were excluded from the analyses because of a large number of trials (greater than 100) with poor pupil signal quality. Remaining trials with poor signal quality, as well as trials in which participants responded incorrectly, were excluded from the analysis (10.9% of the dataset; no difference across conditions: Friedman test, $\chi^2 = 0.05$, $p = 0.82$). All analyses were conducted in a similar way to experiment 1.

(b) Results

Similar to experiment 1, we found that pupil response to the first operand was larger under conditions of large IG ($F_{1,17.4} = 5.17$, $p = 0.036$, β mean (95% CI): 0.0723 (0.0099–0.1346); see figure 3A). Likewise, RTs were also affected by IG condition ($F_{1,16.9} = 22.6$, $p < 0.001$, β mean (95% CI): −0.2132 (−0.301 to −0.125)) and pupil response to first operand ($F_{1,17.2} = 31.9$, $p < 0.001$, β mean (95% CI): −0.1038 (−0.140 to −0.0677); see figure 3B).

(c) Interim discussion

The results from experiment 2 showed that, similar to experiment 1, in conditions in which the first operand was more informative, the eye pupil dilated significantly more and was associated with shorter RTs. These findings confirm our initial interpretation that hearing the first operand leads to updating of the priors over possible responses, in agreement with a Bayesian take on arithmetic problem-solving. In the present experiment, the first operands were identical across conditions, meaning that the effect of the first operand on pupil dilation cannot be explained by low-level phonological or lexical differences.

However, experiment 2 left open an alternative interpretation in terms of anticipatory calculation. Indeed, the rationale of this experiment was to match easy and hard conditions for the first operands but to make them more or less informative by varying the range of possible values taken by the second operands. Because the easy condition allowed only two possible values (+2 or +8), participants may have anticipated the result by computing both alternatives in parallel immediately after hearing the first operand. This strategy would provide them with a processing advantage that could explain the increase in pupil dilation in response to the first operand, compared with the hard condition, regardless of the associated IG. Experiment 3 was therefore motivated by the need to further specify the relationship between the IG and the pupil response to the first operand. We created four different conditions that only differed in the number of possible values that the second operand could take: either 2, 4, 12 or 27 values. This allowed us to look for a linear relationship between pupil dilation and IG across four different levels, including two additional conditions where anticipatory calculation was unlikely because it would require computing up to 12 or 27 additions in parallel.

4. Experiment 3

In the third experiment, we created four independent conditions in which the first operand was always taken from the same list of two-digit numbers, whereas the second operand was a two-digit number taken from a list of varying length (i.e. 2, 4, 12 or 27

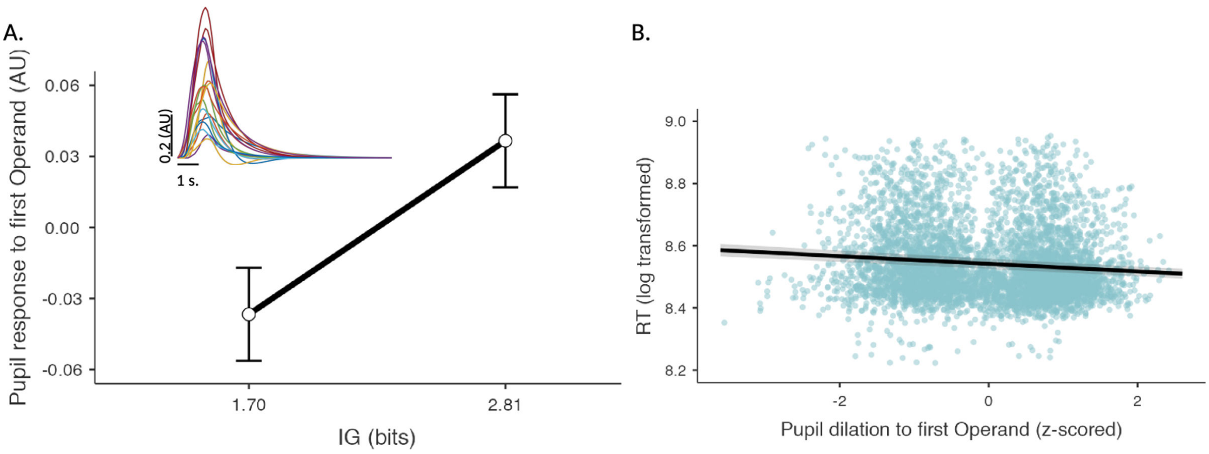

Figure 3. (A) Pupil dilation to first operand presentation as a function of its information gain. (B) RT as a function of pupil dilation to the first operand. Same conventions as in figure 2.

choices), resulting in four different levels of IG. If pupil size reflects cognitive cost linearly related to uncertainty reduction [14], then a linear relationship with IG is predicted.

(a) Method

(i) Participants

Twenty French-speaking participants took part in this experiment (eight males; one left-handed; $M = 23.4$, s.d. = 3). They all had normal or corrected-by-lenses vision and did not report any antecedent of mathematical learning disability when asked by the experimenter. The experiment was in accordance with the ethical standards established by the Declaration of Helsinki and was approved by the local ethics committee.

(ii) Stimulus materials

The first operands consisted of double-digit French number words whose magnitude ranged from 21 to 49, excluding numbers ending in ‘0’ (i.e. 20, 30, 40). The second operands were separated into four different conditions. In the first, most informative condition, they were {‘22’, ‘48’}. In the second, most informative condition, they were {‘22’, ‘28’, ‘42’, ‘48’}. In the third condition, they were {‘22’, ‘23’, ‘24’, ‘26’, ‘27’, ‘28’, ‘42’, ‘43’, ‘44’, ‘46’, ‘47’, ‘48’}. In the fourth condition, they could consist of any number between 21 and 49, with the exclusion of numbers ending in ‘0’. The experiment counted 108 trials for each condition, giving 432 trials in eight separated blocks. In the present experiment, the IG was 3.22 bits for the first condition, 2.53 bits for the second, 1.44 bits for the third and 0.59 bits for the fourth.

(iii) Task and procedure

The procedure was the same as in experiments 1 and 2 except for the longer fixation duration (7 s). This was added in order to further ensure that conditions would not differ in tonic pupil size. The experiment was divided into two different sessions in order to avoid participant fatigue. Each session included one block of each IG condition. Before every block of trials, participants were reminded of the range of the second operand and of the possible responses. The order of the blocks was counterbalanced across all participants, and this order was kept constant between the two sessions.

(iv) Data analysis

Three sessions from different participants were excluded from the analysis owing to the poor quality of the pupil signal (more than 100 trials without usable signal). All analyses were performed similarly to previous experiments, except that the IG associated with the four conditions of arithmetic difficulty was introduced as a continuous variable. Erroneous trials were excluded from the analysis (13% of the dataset; no difference across conditions: Friedman test, $\chi^2 = 1.16$, $p = 0.76$).

In order to test whether IG explains unique variance in pupil dilation, we then aggregated the data from the three experiments and estimated a linear model with pupil betas as the dependent variable, and IG, RT, success expectation (condition-based accuracy) and experiment (1–3) as independent variables. Condition-based accuracy was computed individually to estimate the expectation of being correct as a function of trial type.

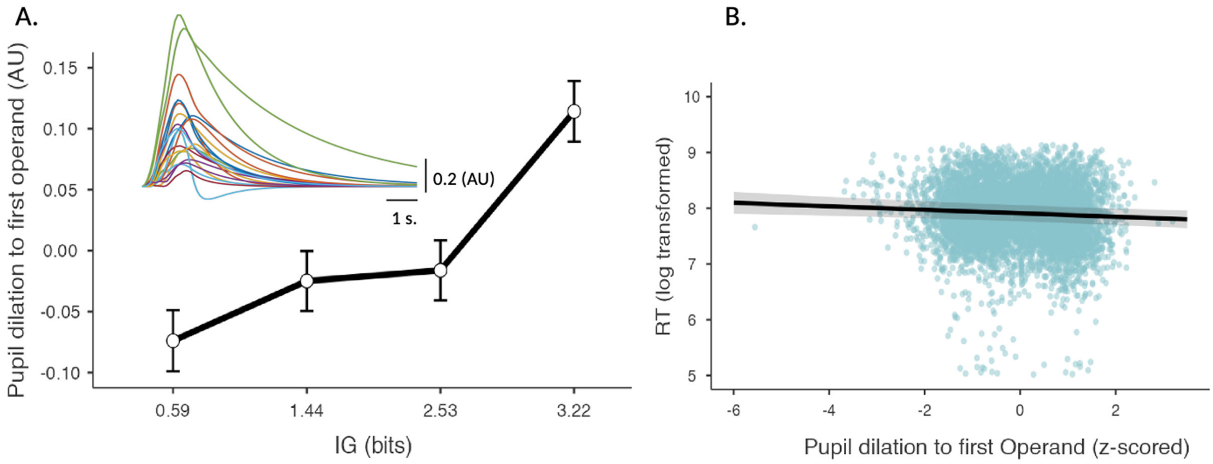

Figure 4. (A) Pupil dilation to first operand presentation as a function of its information gain. (B) RT as a function of pupil dilation to the first operand. Conventions are identical to those of figure 2.

(b) Results

The analysis revealed a significant effect of IG on pupil dilation to the first operand ($F_{1,18.4}$ = 4.68, p = 0.04, β mean (95% CI): 0.0559 (0.0052–0.1066)). Similar to the other experiments, the RT analysis highlighted a significant main effect of IG ($F_{1,19.4}$ = 8.81, p = 0.008, β mean (95% CI): −0.0875 (−0.1452 to −0.0297)) and pupil size ($F_{1,17.3}$ = 12.98, p = 0.002, β mean (95% CI): −0.0572 (−0.0884 to −0.0261); see figure 4B).

In order to further support the hypothesis of a linear relationship between IG and pupil dilation, we aggregated the data of the three experiments and replicated the analysis on pupil dilation with IG as a continuous variable (figure 5). The results confirmed the linear effect of the IG used across the three experiments on pupil dilation to the first operand (using both subjects and experiments as random factors: $F_{1,58}$ = 14.5, p < 0.001, β mean (95% CI): 0.1687 (0.0811–0.256)).

Finally, in order to test whether IG explains unique variance in pupil dilation, we estimated a linear model with pupil betas as the dependent variable, and IG, RT, success expectation (condition-based accuracy) and experiment (1–3) as independent variables. The results showed that the effect of IG on pupil betas remained significant ($F_{1,54.3}$ = 9.77, p < 0.003, β mean (95% CI): 0.0399 (0.01487–0.0649)), after controlling for the variance explained by experiment ($F_{2,21.6}$ = 13.53, p < 0.001), RT ($F_{1,38.5}$ = 42.52, p < 0.001, β mean (95% CI): −0.6698 (−0.87107 to −0.4684)) and success expectation ($F_{1,16}$ = 1.67, p < 0.214, β mean (95% CI): 0.2427 (−0.12517 to 0.6105)).

5. General discussion

The present study aimed to investigate pupil dilation in relation to the updating of prior distributions over possible responses in arithmetic problem-solving. Across the three experiments, we found that the first operand induced a larger pupil dilation in conditions where it carried more information, even after controlling for the variance explained by RT and success expectation. Moreover, in all experiments, this increase in pupil size was negatively correlated with RT, showing that the informational value of the first operand, reflected in pupil dilation, helped participants solve the arithmetic problem. These findings corroborate the hypothesis that arithmetic problem-solving relies on progressive updating of prior distribution over possible responses, in agreement with Bayesian models of behaviour [1,34] and that IG can be used as measures of the cognitive cost of arithmetic problem-solving and is a decisive factor of its perceived difficulty [14].

While our findings support a Bayesian interpretation of arithmetic problem-solving, several limitations should be acknowledged. First, we cannot entirely rule out heuristic or subset-based processing strategies, such as participants mentally sampling a few plausible sums rather than representing the full distribution. However, the graded effect of IG across the four conditions in experiment 3, including operand sets too large for exhaustive evaluation (e.g. 27 possible second operands), strongly suggests that participants were engaged in probabilistic updating rather than discrete enumeration.

Second, a potential concern is that the observed effects of IG on pupil dilation and RTs may reflect variations in WM load rather than Bayesian updating *per se* [35]. Our experimental design directly addressed this possibility: across all three experiments, IG and WM load varied inversely, with higher IG conditions corresponding to situations where the first operand allowed participants to eliminate more candidate outcomes, thereby reducing the number of plausible results that needed to be retained or manipulated in WM. In contrast, low IG conditions preserved a wider range of outcome possibilities, increasing uncertainty and expanding the memory space required to support ongoing calculations. This inverse relationship suggests that the observed increase in pupil dilation in high IG conditions is unlikely to be owing to increased WM load alone. Our findings, instead, align with the interpretation that pupil size reflects the cost of integrating information to update priors over possible outcomes—a hallmark of Bayesian inference—rather than raw memory storage demands.

A third concern regarding the present study is that, although we interpret pupil dilation as reflecting IG, this measure may also be modulated by other cognitive or affective factors such as emotional responses, which we cannot fully dissociate from the

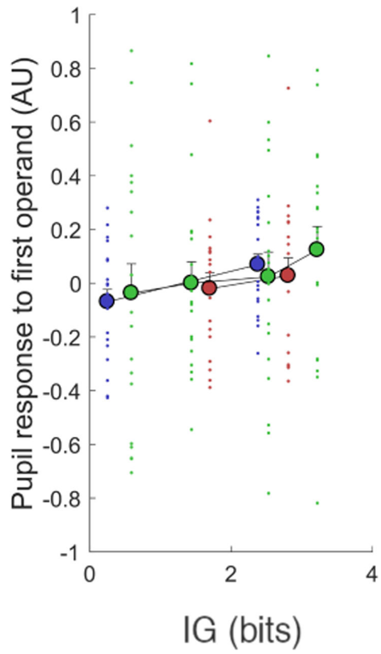

Figure 5. Pupil dilation as a function of information gain across all three experiments. Individual dots correspond to individual subjects (blue: experiment 1; red: experiment 2; green: experiment 3).

updating process. Future studies using converging behavioural and neural measures could help disentangle these components and test whether Bayesian updating operates jointly with heuristic shortcuts during symbolic computation.

This conclusion does not contradict previous accounts of arithmetic problem-solving, which have highlighted the effect of problem size [36,37] or carry operations on associated workload [30]. On the contrary, the perspective we defend in this work could help explain some of these effects. The problem size effect refers to the decreased performance observed when individuals have to process larger operands in arithmetic problems and has classically been attributed to frequency effects, larger numbers being encountered less frequently than smaller ones [36]. The Bayesian perspective used here is in agreement with this view, since more frequent encounters of the operands, together with their sum and difference, would favour their representation in the prior probability distributions, thereby improving performance [14]. It is also compatible with the idea that the representation of large numbers has more uncertainty [38,39], which would translate in Bayesian terms to wider, less peaked priors. Updating from such uncertain priors requires more information (higher entropy reduction), which maps onto greater IG and greater processing effort.

On the other hand, the carry effect refers to the finding that performance decreases with the number and the value of the carries that need to be maintained and added in the leftward column in multi-digit calculation (e.g. 246 + 381). The WM load incurred by carry operations is not so high, as evidenced by the low impact of articulatory suppression, an experimental manipulation that prevents verbal rehearsal [30,40]. Yet, executive interference has been observed when the calculation imposes a sequence of carries of increasing value. Such interference has been explained by the competition between two task sequences, one without and one with the execution of a carry [40]. This explanation can be subsumed under the Bayesian perspective proposed here. Carry operations introduce additional uncertainty in the internal representation of intermediate outcomes, requiring updates not just to digit-level priors but to a latent variable indicating whether a carry is needed. This effectively increases the entropy of the posterior distribution and necessitates a larger IG to converge on a final response. As such, carry operations can be conceptualized as hidden variables within a probabilistic model of arithmetic, and the cognitive cost they impose is reflected in greater pupil dilation and longer RTs.

The grounding of arithmetic problem-solving in an information processing framework also provides a new perspective on the interaction between exact and approximate arithmetic processes [41–44]. We propose that approximate arithmetic would allow individuals to perform an initial, fast shaping of the prior distributions over responses, which would then be used as a starting point for exact arithmetic computations. Such combinations of approximate and precise processes have been shown to be advantageous from the information-theoretic perspective [45]. From a more mechanical point of view, this initial, rapid but approximate shaping of prior probability distributions over possible responses could reduce the complexity of the WM manipulations by eliminating unlikely options. We recently found a trace of this heuristic process in the eye movements performed by adults and children solving addition and subtraction problems [46,47]. We used lateralized eye movements to test the hypothesis that addition and subtraction shift attention in opposite directions along a left-to-right oriented continuum representing numbers in ascending order. A time-course analysis revealed that addition deviated the gaze rightward, compared with subtraction, as soon as participants had processed the first operand and the operator. The predictive nature of these attention shifts, which typically occur before the presentation of the second operand, clearly shows that their role is not to identify the exact answer but to give an approximate idea of plausible answers that would then be compared with the solution obtained by an exact procedure. We proposed a probabilistic model where attention shifts are embedded in WM processes: spatial resources would be used to adjust the range of plausible answers as the calculation progresses, thereby reducing the part of the number sequence maintained in verbal WM for precise incrementation. The present study provides direct empirical support to this probabilistic model of arithmetic problem-solving, highlighting the need to consider prediction as a core determinant of the interactions between space and number [48].

Finally, the Bayesian framework proposed here also resonates with probabilistic interpretations of number comparison. Prior work has shown that the symbolic size effect depends on the frequency with which specific numbers occur, while the distance effect depends on the conditional probability that one number is larger or smaller than another [49]. Within our framework, both effects can be seen as manifestations of prior shaping through long-term experience: frequent numbers acquire stronger priors, which facilitate their processing, whereas numbers more often encountered as 'larger' evoke stronger prior responses associated with that concept. Thus, frequency- and probability-based accounts of symbolic number comparison naturally converge with our Bayesian interpretation of arithmetic problem-solving, in which performance emerges from the interaction between learnt priors and the conditional probabilities linking stimuli and responses. This correspondence supports the view that both symbolic comparison and arithmetic computation rely on common probabilistic mechanisms governing the formation and updating of numerical representations.

In summary, the present work proposes a new viewpoint on mental arithmetic processes and their associated cognitive cost, based on the principle of Bayesian inference. Empirical validation of the proposed framework shows that Bayesian inference provides a theoretically valid explanation of human behaviour, which is not limited to basic sensorimotor functions but encompasses more advanced cognitive functions learnt through cultural practice, such as symbolic arithmetic. At a more mechanistic level, our probabilistic model explains how approximate arithmetic can help refine the range of plausible responses in advance of defining exact arithmetic procedures, thereby alleviating the WM load. Future research is needed to enlarge this framework to other dimensions of numerical cognition.

Ethics. The experiment was in accordance with the ethical standards established by the Declaration of Helsinki and was given prior approval by the local ethics committee (EudraCT: 2018-A01233-52, CPP: 18-INSB-01).

Data accessibility. All data and statistical models are available at https://researchbox.org/3762.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. A.Z.: conceptualization, formal analysis, funding acquisition, methodology, project administration, resources, software, supervision, visualization, writing—original draft; S.S.: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft; M.A.: conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. The research was supported by grant T.0006.20 from the FRS-FNRS and grant ARC21/26-112 from the Fédération Wallonie Bruxelles (FWB, Belgium) as well as ANR VICONTE 2018-0524 and ANR CoCogIT 2019-0054.

Acknowledgements. A.Z. is a research associate at the Centre National pour la Recherche Scientifique (CNRS, France). M.A. is a research associate at the Fonds National de la Recherche Scientifique (FRS-FNRS, Belgium).

References

1. Colombo M, Seriès P. 2012 Bayes in the brain—on Bayesian modelling in neuroscience. *Br. J. Phil. Sci.* **63**, 697–723. (doi:10.1093/bjps/axr043)
2. Knill DC, Pouget A. 2004 The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719. (doi:10.1016/j.tins.2004.10.007)
3. de Lange FP, Heilbron M, Kok P. 2018 How do expectations shape perception? *Trends Cogn. Sci.* **22**, 764–779. (doi:10.1016/j.tics.2018.06.002)
4. Körding KP, Wolpert DM. 2004 Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247. (doi:10.1038/nature02169)
5. Xu F, Tenenbaum JB. 2007 Word learning as Bayesian inference. *Psychol. Rev.* **114**, 245–272. (doi:10.1037/0033-295X.114.2.245)
6. Ashcraft MH. 1987 Children's knowledge of simple arithmetic: a developmental model and simulation. In *Formal methods in developmental psychology* (eds J Bisanz, CJ Brainerd, R Kail), pp. 302–338. New York, NY: Springer New York. (doi:10.1007/978-1-4612-4694-7_9)
7. Siegler RS. 1988 Strategy choice procedures and the development of multiplication skill. *J. Exp. Psychol.: Gen.* **117**, 258–275. (doi:10.1037//0096-3445.117.3.258)
8. Rinaldi L, Parente L, Marelli M. 2022 Toward a unified account of nonsymbolic and symbolic representations of number: insights from a combined psychophysical-computational approach. *Psychon. Bull. Rev.* **29**, 985–994. (doi:10.3758/s13423-021-02043-3)
9. De Visscher A, Noël MP. 2014 The detrimental effect of interference in multiplication facts storing: typical development and individual differences. *J. Exp. Psychol.: Gen.* **143**, 2380–2400. (doi:10.1037/xge0000029)
10. Hamann MS, Ashcraft MH. 1986 Textbook presentations of the basic addition facts. *Cogn. Instr.* **3**, 173–202. (doi:10.1207/s1532690xci0303_2)
11. Ashcraft MH. 1992 Cognitive arithmetic: a review of data and theory. *Cognition* **44**, 75–106. (doi:10.1016/0010-0277(92)90051-i)
12. Zénon A, Solopchuk O, Pezzulo G. 2019 An information-theoretic perspective on the costs of cognition. *Neuropsychologia* **123**, 5–18. (doi:10.1016/j.neuropsychologia.2018.09.013)
13. Zénon A. 2019 Eye pupil signals information gain. *Proc. R. Soc. B* **286**, 20191593. (doi:10.1098/rspb.2019.1593)
14. Alamia A, VanRullen R, Pasqualotto E, Mouraux A, Zenon A. 2019 Pupil-linked arousal responds to unconscious surprisal. *J. Neurosci.* **39**, 5369–5376. (doi:10.1523/JNEUROSCI.3010-18.2019)
15. Bradley MM, Miccoli L, Escrig MA, Lang PJ. 2008 The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607. (doi:10.1111/j.1469-8986.2008.00654.x)
16. van der Wel P, van Steenbergen H. 2018 Pupil dilation as an index of effort in cognitive control tasks: a review. *Psychon. Bull. Rev.* **25**, 2005–2015. (doi:10.3758/s13423-018-1432-y)
17. Hess EH, Polt JM. 1964 Pupil size in relation to mental activity during simple problem-solving. *Science* **143**, 1190–1192. (doi:10.1126/science.143.3611.1190)
18. Joshi S, Li Y, Kalwani RM, Gold JI. 2016 Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* **89**, 221–234. (doi:10.1016/j.neuron.2015.11.028)
19. Salvi V, Courtand G, de Deurwaerdère P, Cardoit L, Valerio S, Delcasso S, Georges F, Michelet T. 2025 Cingulate cortex stimulation drives distinct pupillary responses in rat via recruitment of noradrenergic neurons in the locus coeruleus. *Cereb. Cortex* **35**, bhaf085. (doi:10.1093/cercor/bhaf085)

20. Preuschoff K, 't Hart BM, Einhäuser W. 2011 Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front. Neurosci.* **5**, 115. (doi:10.3389/fnins.2011.00115)
21. Vincent P, Parr T, Benrimoh D, Friston KJ. 2019 With an eye on uncertainty: modelling pupillary responses to environmental volatility. *PLoS Comput. Biol.* **15**, e1007126. (doi:10.1371/journal.pcbi.1007126)
22. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI. 2012 Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–1046. (doi:10.1038/nn.3130)
23. Silvetti M, Vassena E, Abrahamse E, Verguts T. 2018 Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLoS Comput. Biol.* **14**, e1006370. (doi:10.1371/journal.pcbi.1006370)
24. Zénon A. 2017 Time-domain analysis for extracting fast-paced pupil responses. *Sci. Rep.* **7**, 41484. (doi:10.1038/srep41484)
25. Cochrane D, Orcutt GH. 1949 Application of least squares regression to relationships containing auto-correlated error terms. *J. Am. Stat. Assoc.* **44**, 32–61. (doi:10.1080/01621459.1949.10483290)
26. Chen Y, Campbell JID. 2018 'Compacted' procedures for adults' simple addition: a review and critique of the evidence. *Psychon. Bull. Rev.* **25**, 739–753. (doi:10.3758/s13423-017-1328-2)
27. Fayol M, Thevenot C. 2012 The use of procedural knowledge in simple addition and subtraction problems. *Cognition* **123**, 392–403. (doi:10.1016/j.cognition.2012.02.008)
28. Campbell JI, Xue Q. 2001 Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* **130**, 299–315. (doi:10.1037//0096-3445.130.2.299)
29. Nuerk HC, Moeller K, Willmes K, Fischer MH. 2011 Extending the mental number line: a review of multi-digit number processing. *Zeitschrift Für Psychologie* **219**, 1–2. (doi:10.1027/2151-2604/a000040)
30. Fürst AJ, Hitch GJ. 2000 Separate roles for executive and phonological components of working memory in mental arithmetic. *Mem. Cogn.* **28**, 774–782. (doi:10.3758/bf03198412)
31. Klein E, Moeller K, Dressel K, Domahs F, Wood G, Willmes K, Nuerk HC. 2010 To carry or not to carry--is this the question? Disentangling the carry effect in multi-digit addition. *Acta Psychol.* **135**, 67–76. (doi:10.1016/j.actpsy.2010.06.002)
32. Moeller K, Klein E, Nuerk HC. 2011 Three processes underlying the carry effect in addition – evidence from eye tracking. *Br. J. Psychol.* **102**, 623–645. (doi:10.1111/j.2044-8295.2011.02034.x)
33. Seitz K, Schumann-Hengsteler R. 2022 Phonological loop and central executive processes in mental addition and multiplication. *Psychol. Test Assess. Model.* **44**, 275.
34. Penny W. 2012 Bayesian models of brain and behaviour. *ISRN Biomath.* **2012**, 1–19. (doi:10.5402/2012/785791)
35. Bays PM, Husain M. 2008 Dynamic shifts of limited working memory resources in human vision. *Science* **321**, 851–854. (doi:10.1126/science.1158023)
36. Campbell JI. 2005 *The handbook of mathematical cognition*. Psychology Press. See https://books.google.com/books?hl=fr&lr=&id=8Vx6OSC8ffkC&oi=fnd&pg=PR3&dq=campbell+2005+problem+size&ots=H_0azz-_rS&sig=ytFHyZQmwgxddDIqVCIEh4q0oo.
37. Verguts T, Fias W. 2005 Neighbourhood effects in mental arithmetic. *Psychol. Sci.* **47**, 132.
38. Dehaene S. 2003 The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* **7**, 145–147. (doi:10.1016/s1364-6613(03)00055-x)
39. Whalen J, Gallistel CR, Gelman R. 1999 Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* **10**, 130–137. (doi:10.1111/1467-9280.00120)
40. Imbo I, Vandierendonck A, De Rammelaere S. 2007 The role of working memory in the carry operation of mental arithmetic: number and value of the carry. *Q. J. Exp. Psychol.* **60**, 708–731. (doi:10.1080/17470210600762447)
41. Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S. 1999 Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* **284**, 970–974. (doi:10.1126/science.284.5416.970)
42. Hyde DC, Khanum S, Spelke ES. 2014 Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* **131**, 92–107. (doi:10.1016/j.cognition.2013.12.007)
43. Park J, Brannon EM. 2014 Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition* **133**, 188–200. (doi:10.1016/j.cognition.2014.06.011)
44. Sekeris E, Verschaffel L, Luwel K. 2021 Exact arithmetic, computational estimation and approximate arithmetic are different skills: evidence from a study with 5-year-olds. *Infant Child Dev.* **30**, e2248. (doi:10.1002/icd.2248)
45. Genewein T, Leibfried F, Grau-Moya J, Braun DA. 2015 Bounded rationality, abstraction, and hierarchical decision-making: an information-theoretic optimality principle. *Front. Robot. AI* **2**, 27. (doi:10.3389/frobt.2015.00027)
46. Masson N, Dormal V, Stephany M, Schiltz C. 2024 Eye movements reveal that young school children shift attention when solving additions and subtractions. *Dev. Sci.* **27**, e13452. (doi:10.1111/desc.13452)
47. Salvaggio S, Masson N, Zénon A, Andres M. 2022 The predictive role of eye movements in mental arithmetic. *Exp. Brain Res.* **240**, 1331–1340. (doi:10.1007/s00221-022-06329-3)
48. Hawes Z, Ansari D. 2020 What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychon. Bull. Rev.* **27**, 465–482. (doi:10.3758/s13423-019-01694-7)
49. Krajcsi A, Kojouharova P, Lengyel G. 2022 Processing symbolic numbers: the example of distance and size effects. In *A life in cognition. language, cognition, and mind* (eds J Gervain, G Csibra, K Kovács), pp. 379–394, vol. **11**. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-030-66175-5_27)