Article:

MaBouDi, H. orcid.org/0000-0002-7612-6465, Barron, A.B., Li, S. et al. (8 more authors) (2021) Non-numerical strategies used by bees to solve numerical cognition tasks. Proceedings of the Royal Society B: Biological Sciences, 288 (1945). 20202711. ISSN 0962-8452

https://doi.org/10.1098/rspb.2020.2711

Non-numerical strategies used by bees to solve numerical cognition tasks

HaDi MaBouDi[1,#], Andrew B Barron[1,2,#], Sun Li[3], Maria Honkanen[4], Olli Loukola[4], Fei Peng[3], Wenfeng Li[5], James A. R. Marshall[1], Alex Cope[1], Eleni Vasilaki[1], and Cwyn Solvi[2,6,*]

[1] Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK
[2] Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, AU
[3] Department of Psychology, School of Public Health, Southern Medical University, Guangzhou, China
[4] Department of Ecology and Genetics, University of Oulu, Oulu, Finland
[5] Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Science, Guangzhou, China
[6] School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK
[#] authors contributed equally to this work
* corresponding author: c.solvi@qmul.ac.uk

Abstract

If we are truly interested in the evolution of cognition in its many forms, it is vital to understand not simply that an animal can solve a task, but how they solve the task. We examined how bees solve a numeric-based task utilizing stimuli common amongst numerical cognition studies. Bees performed well on the task, but additional tests showed that they had learned continuous (non-numerical) cues. In simulations, a simple network model containing just nine elements was capable of learning the task using only continuous cues inherent in the training stimuli, with no numerical processing. This model was able to reproduce behaviours that have been considered in other studies indicative of numerical cognition. Our results support the idea that a sense of magnitude may be more primitive and basic than a sense of number. Our findings highlight how problematic inadvertent continuous cues can be for studies of numerical cognition. This remains a deep issue within the field that requires increased vigilance and cleverness from the experimenter. We suggest ways of better assessing numerical cognition in non-speaking animals, including assessing the use of all alternative cues in one test, using cross-modal cues, analysing behavioural responses to detect underlying strategies, and finding the neural substrate.

1. Introduction

Mapping specific cognitive capacities to the behaviour of any animal is rarely straightforward. The difficulty is that animals may not be solving the task the way we think they are. One example of this is in our own recent work where we had bees discriminate different shapes based on relative size [1]. Bees' performance increased over training to well above chance, and in the unrewarded test they seemed to have learned to discriminate shapes based on relative size. However, analysis of first and sequential choices during training bouts and tests revealed that the bees actually switched to a simpler strategy in the middle of training: win-stay/lose-switch. These results, along with other works suggesting animals are able to solve tasks in unexpected ways (e.g. [2–7]), prompted us to look deeper into the strategies of animals in numerical cognition tasks.

Numerical cognition has been claimed in a large number of animal species (e.g. [8–39]), suggesting that a sense of number is widespread (for reviews see [40–42]). By far, the most common method for testing numerical cognition in non-verbal animals is to have subjects discriminate 2D visual displays with differing numbers of shapes (Fig 1; [8–39] all used this design). As pointed out by others (e.g. [43,44]), in these types of designs, continuous (non-numerical) cues often unavoidably covary with numerosity. These include size and shape of elements, area (total amount of colour), edge length (total boundary length of elements), convex hull (the minimum convex region covering all elements, spatial frequency (the amount of alternating dark and light regions), and illusory contour (the basic shape that outlines all elements). In Figure 1 we further describe these cues and their natural covariation with number (figure *1a*-*e*). This covariation makes it difficult to know whether animals actually used any sense of number to solve their tasks.

The issue of non-numerical strategies within numerical cognition studies has been highlighted by others [43–46]. It was established decades ago that cells within the visual system respond to various continuous visual features [45,47,48] and it has long been known that continuous features can be reliable discrimination cues, even for bees [49–52]. Further, several works show that animals use non-numerical cues to solve numeric-based tasks when not controlled for, e.g. size of elements [53], total area [54], and convex hull [55], and even when they are controlled (e.g. [56]; see Discussion).

Most studies investigating numerical cognition attempt to control for at least one non-numerical cue. Several works have made valiant efforts to control for most continuous cues (e.g. [57,58]). However, we have found no studies that tested for all continuous variables. It seems clear that animals are solving these tasks, but the question we attempt to address here is how they might be solving the tasks. We set out to determine how honeybees solve a numeric-based task using stimuli common amongst numerical cognition studies.

2. Materials and Methods

(a) Subjects

Honeybees (*Apis mellifera*) used in the experiment were maintained at the University of Oulu (Oulu, Finland) and at Guangdong Institute of Applied Biological Resources (Guangzhou, China) in September and November 2019, respectively. Prior to training, honeybees were fed ad libitum from a gravity feeder providing 30% sucrose solution. Each focal honeybee was first lured to visit the experimental setup by allowing her to drink and walk onto a cotton bud soaked in 50% sucrose solution and then transferring the bee to the setup. Each forager that returned to the setup on her own was marked on her thorax with a coloured dot for identification.

(b) Experimental setup and procedure

The setup consisted of a 50 x 50 cm acrylic sheet. Stimuli were 6 x 6 cm white displays (laminated sheets of paper) with between 1 and 4 black shapes (squares, diamonds, or circles). The stimuli, identical to those used in [28], were presented vertically with a landing platform attached just below the displays. Stimuli were randomly allocated for each bee and changed when the bee returned to the hive to offload sucrose, prior to her returning to the setup. The spatial arrangement of stimuli could be randomly changed, thus excluding position orientation cues. The background acrylic sheet and landing platforms were grey coloured. The acrylic background sheet, hangers, platforms and displays were washed with water then 70 % ethanol between all visits to exclude the use of olfactory cues. Two shapes were used in training, and the third shape was used for testing. During training, honeybees found either a 10μL droplet of 50 % sucrose solution or a 60 mM quinine hemisulphate solution, for correct and incorrect choices, respectively. Each trial, four stimuli (two identical correct; two identical incorrect) were presented simultaneously on the acrylic sheet. Stimuli positions were changed after each choice to new random positions. A choice was defined as any time a honeybee landed on a platform and touched the solution (sucrose or quinine) with their proboscis, leg, or antenna.

One group of bees (n = 10) was trained to associate stimuli consisting of more elements with a reward, and a second group of bees (n = 10) was trained to associate stimuli consisting of fewer elements with a reward. The choices of individual bees during training were recorded until a criterion of ≥ 80 % for any 10 consecutive choices was reached (after a minimum of 20 conditioned choices). Once an individual bee reached criterion, she was presented with a learning test followed by two additional control tests. Bees reached criterion on average in 41 ± 8 choices. Each test lasted two minutes and all choices made were recorded as the dependent variable for statistical analyses. During all tests, a 10μl of unrewarding water was placed on each platform. Between tests, bees received two reinforced refresher trials (with the same stimuli used in training) to maintain motivation. For the learning test, bees were presented with the shape that they had not been trained on – the purpose being to test whether bees learned to solve the task. The two control tests examined whether honeybees used the number of elements or continuous visual cues. The first control test (Equal/Incongruent test) had two pairs of stimuli, each with two elements, but one pair with higher edge length, convex hull, and spatial frequency. The second control test (Incongruent/Opposite test) also had two pairs of stimuli, one pair with three elements and the other with two elements but still with higher edge length, convex hull, and spatial frequency. In all tests, total black surface area was the same across all stimuli. Experiments were performed by three different groups of individuals (MH and OL, SL, and CS) to help independently verify the results.

(c) Statistical analyses

R 3.6.1 with library “lme4” was used to perform all generalised linear mixed-effect models (glmm) with binomial distribution and logit function. For the glmm evaluating the results of the tests, country and rule (more-than/less-than) were considered as fixed factors and bee ID as a random effect (Table S1). Because country and rule had no effect on performance, we display data as the mean ± s.e.m. of all bees' data. We then removed country and rule in a second glmm (Table 2). Our second model ranked better than the first on the grounds of Akaike's Information Criterion [59] adjusted for small sample sizes (AICc), and therefore we present data from this second model in the main text. For analyses of all test videos, a blind protocol was employed, in that each video filename was coded so that the experimenter doing the analysis was blind to the training of each bee.

To calculate the spatial frequency of the training and test stimuli, a two-dimensional Fourier transform on each image was performed, followed by a power spectrum calculation as the square amplitude of the Fourier transform and averaged over orientation [60]. The actual power over all frequencies was then measured by calculating the area under the curve of the radially averaged power spectrum. Calculations for spatial frequency, convex hull and edge length were done in MATLAB 2018b (MathWorks, Mass., USA). Statistical analyses for the model results were also performed in MATLAB 2018b.

(d) Neural network model

Our model utilizes spatial frequency encoding that is supported by bees' ability to discriminate visual patterns based on spatial frequency [49,50] and observed neurons in the visual lobe of insects that provide a mechanism of frequency coding [61,62]. Analogous to the spatial frequency coding in primates [63,64], bees may use Gabor-like filters in their visual lobe to extract spatial frequency information from visual stimuli [65]. For our model, the stimulus, s, is encoded by the activity of a population of neurons with different preferred spatial frequency that possess similar response profiles. The evoked spiking activities of the seven sensory neurons were simulated by fixed Gaussian tuning curves spanning spatial frequencies of the input from zero to six as $g_i\ (s, \sigma) = R_0 + R_{Max}\ exp[-\frac{1}{2\sigma^2}(s - f_i\ )^2\ ] + \aleph(0, \sigma_N)$, where $R_0 = 50\ \text{spike/sec}$ and $R_{Max} = 200\ \text{spike/sec}$ are the spontaneous and maximum firing-rate of the sensory neuron. $\sigma = 2.5$ controls the degree of the selectivity of the sensory neurons to different frequencies around the preferred frequency, f_i. Gaussian noise, $\aleph(0, \sigma_N)$ model the randomness of neural activities.

Outputs of all sensory neurons drive a decision neuron through a vector of synaptic weights, W, to create the decision neuron's activity in response to the input, as:

$$D(s) = F\left(\sum_{k=0}^{6} W_k . g_k(s, \sigma)\, ; a, b\right)$$

where $F(x; a, b) = A_0/(1 + exp(-a(x - b))\ )$ is the activation function with the maximum activity at $A_0 = 100$ Spike/sec. The parameters $a = 0.05$ and $b = 50$ control the sensitivity of the neuron to the input and spontaneous activity of the decision-neuron, respectively.

Since we assume that the difference of the decision neuron's responses to the positive (s_p) and negative stimuli (s_n) must be increased during the training phase, the locally optimal synaptic weights, W^{opt}, can be obtained from maximizing the objective function:

$$L = \sum_{t=1}^{m} [D(s_p^t) - D(s_n^t)]\ r^t,$$

where t and m are the index over the paired stimuli and the number of presented stimuli, respectively. Here, r presents the reinforcement signal (VUM-mx1 neuron) that provides modulated feedback whether a stimulus is paired with the reward or punishment ($r = 1$) and $r = 0$ for when no reinforcement signals is presented. The (on-line) updates of the synaptic weights, W_i^t are calculated by

$$W_i^t = W_i^{t-1} + \eta \, \frac{\partial}{\partial W_i}\left(D\left(s_p^t\right) - D\left(s_n^t\right)\right) r^t$$

where η is the rate of the weights change. W_i^{t-1} is the updated weight from the iteration $t-1$ (with W_i^0 being the initial weight), and

$$\frac{\partial}{\partial W_i}\left(D\left(s_p^t\right) - D\left(s_n^t\right)\right)$$

$$= g_i\left(s_p^t, \sigma\right) F'\left(\sum_{k=0}^{6} W_k . g_k\left(s_p^t, \sigma\right) ; a, b\right) - g_i\left(s_n^t, \sigma\right) . F'\left(\sum_{k=0}^{6} W_k . g_k\left(s_n^t, \sigma\right) ; a, b\right)$$

Finally, the derivatives of the activation function F is obtained as

$$F'(x; a\, , b) = \frac{A_0 \; a \; exp\left(-a(x-b)\right)}{\left(1 + exp\left(-a(x-b)\right)\right)^2}$$

After exposing the model to conditioned stimuli in learning paradigms, the behavioural outcomes of the model presented with a pair of the test stimuli were evaluated as a simple subtraction of the decision neuron's responses to both test stimuli.

3. Results

(a) Bees use continuous cues over numerosity in a numerical cognition task

Using the same 2D visual stimulus set as a paradigmatic honeybee study [28], and similar to stimulus sets used for other animals (e.g. [8–39]), we first asked whether honeybees use numerosity to solve a numeric-based discrimination task. In this particular stimulus set, area (total black within each stimulus) is kept constant across all stimuli, and therefore could not be used to solve the task. But, similar to many other numerical cognition studies, edge-length (Spearman correlation: rho=0.93, p=1.00e-40), convex hull (Spearman correlation: rho=0.44, p=4.88e-6), and spatial frequency (Spearman correlation: rho=0.92, p=1.00e-40) covaried with number (figure *1f-j*). We therefore aimed to train bees on this stimulus set, for which they've already been shown to discriminate, and subsequently test bees to determine whether they had used these particular continuous cues or rather numerosity to solve the task.

We first trained honeybees (n = 10) to find rewarding sugar solution on displays with more shapes and an aversive quinine solution on displays with fewer shapes (Methods; figure *2a*). Another group of bees (n = 10) was trained on the opposite contingency. Once bees reached 80% performance (8/10 consecutive choices correct), they were given an unrewarded learning test. Bees trained on a “more-than” rule preferred (landed on more often) stimuli containing more elements during the test, whereas bees trained to “less-than” preferred stimuli with fewer elements. Honeybees showed high performance in the Learning test (figure 2*b* left; Generalised linear mixed-effect model (glmm): 95% Confidence interval (CI) = 0.75 (0.47 to 1.03), n = 20, p = 1.49e-07).

To determine if bees used non-numerical cues, after the learning test and refresher trials (Methods), we tested the same honeybees on an “Equal/Incongruent test”, where two pairs of unrewarded stimuli contained the same number of elements (figure 2*b* middle), but differed in edge-length, convex hull, and spatial frequency (figure 2*c-f*). If honeybees were using numerosity, they should prefer all displays equally during this test. Conversely, honeybees more often chose stimuli with a higher level of continuous variables if they had been trained to choose stimuli with more elements, and more often chose stimuli with a lower level of continuous variables if they had been trained to choose stimuli with fewer elements (figure 2*b* middle; glmm: 95% CI = -0.64 (-0.89 to -0.39), n = 20, p = 6.5e-07). This suggests honey bees responded to continuous cues in the stimuli and not the number of elements.

We further tested honeybees on an “Incongruent/Opposite test” where the number of elements for each pair of displays differed (2v3; figure 2*b* right) and the continuous cues (edge length, convex hull, and spatial frequency) were in the opposite direction to the numerical difference (i.e. higher for two elements than for three elements; figure 2*c-f*). In this test, honeybees behaved in the reverse manner to which we would expect if they had learned numerosity. Bees that were trained to associate more elements with reward preferred test displays with the higher level of continuous variables but fewer elements. Bees that were trained to associate fewer elements with reward preferred test displays with the lower level of continuous cues but more elements (figure 2*b* right; glmm: 95% CI = -0.55 (-0.79 to -0.30), n = 20, p = 1.17e-05).

Our results indicate that honeybees use continuous properties to discriminate stimuli with varying number of shapes. This caveat may also apply to other numerical cognition studies with honeybees and other animals that used stimulus sets which controlled for one or more but not all continuous variables.

(b) A neural network model with no reference to numerosity can reproduce behaviors indicative of numerical cognition

Our results beg the question: what explanation is simpler and more plausible: numerical or non-numerical processing? Therefore, how simple is learning continuous variables as an explanation for the behaviour of honeybees? To explore this, we created a simple neural network model containing just nine elements arranged in three layers (figure 3*a*) to encode a relational rule (“more-than” or “less-than”) based only on one non-numerical cue (Materials and Methods). Seven elements acted as sensory neurons that encoded spatial frequency in the visual lobe and which projected frequency information to the eighth element, a single decision neuron in the mushroom bodies (high-level sensory integration centres involved in learning and memory). Synaptic weights between the sensory neurons and decision neuron were adjusted according to the activation (by presentation of stimuli) of the ninth element, a reinforcement neuron, based on the specific learning rule (more-than or less-than). We chose spatial frequency for simplicity, and because we have yet to find any recent study that controlled/tested for it, but the model could also be applied to other continuous variables.

We trained our model following the methods for several experiments in [28], a recent study that had honeybees discriminate 2D visual cues with different numbers of shapes. We then evaluated the model's choices when presented with test stimuli (See Methods for details and figure 3 for simplification). This simple model was able to reproduce the behaviour of honeybees in numerical cognition tasks, with a very simple computational structure using only non-numerical information. Specifically, the model could transfer a “more-than” or “less-than” rule to novel shapes, to stimuli containing a number of elements outside the range trained on, to stimuli with zero elements, and could recognise stimuli with zero elements as the lower end of a continuum (figure 3*b*-*e*). Thus, we are able to reproduce behavioural evidence that has been taken in honeybees (and similarly in other animals) as indicative of understanding number with a model in which there is no processing of numerosity.

4. Discussion

(a) General summary

We are not suggesting that all numerical cognition studies are wrong or that no animal has numerical cognition. We show, however, that in a task using a 2D visual display set with differing number of shapes, non-numerical cues can be learned, they dominate over numerosity when equal to or set in opposition to number of elements, and they can be learned by simple computational systems with no reference to numerosity. Our behavioural and computational results provide a counterexample against the assumption that 2D visual stimuli with different numbers of shapes are processed by honeybees as discrete numerical elements. Our findings suggest that an alternative non-numerical explanation exists for studies using similar methods in honeybees. If other animals are sensitive to any available continuous cues, then an alternative non-numerical explanation exists for those results as well. This is vital information if we truly want to know how any animal solves the numerical problems they face in their own ecological niches.

(b) The depth of the issue

It is very difficult to control for all continuous visual cues [44,66]. By controlling one parameter, another will necessarily covary with numerosity. Even varying parameters randomly during training is not enough to solve the issue. Leibovich and Henik (2014) trained adult humans on visual stimuli of differing number of dots where continuous cues were minimally correlated or uncorrelated with numerosity. Despite this, they found that in a regression analysis, half of the behavioural variance could be explained by the irrelevant continuous cues [56]. Presenting stimuli separately/sequentially may make the task more difficult (e.g. [67,68]). However, animals may store, in working memory, an accumulation of neural responses to continuous variable changes as they pass/observe stimuli, without reference to numerosity [69–72].

It will also not suffice to test for continuous cues separately because animals may learn multiple redundant cues and use those available when others are not [73–78]. Testing all continuous variables *(that cannot be kept constant across stimuli)* and numerosity within one test can help determine if continuous variables have been learned. In one of our recent works, examining how bumblebees solved a numeric-based task, we assessed the use of continuous cues within one unrewarded test [79]. Here, bees were shown 10 stimuli simultaneously during one unrewarded test, each with different numbers of elements and levels of continuous cues. We chose the characteristics of different stimuli so that the bees' choices for some over others would reveal whether or not they had learned and used specific continuous cues to solve the task. For example, two displays both contained the same number of elements, but the elements in one of the displays had a greater edge-length. Bees chose these two displays equally in the test, suggesting they did not use edge length. However, if they had performed well on the test (i.e. more often chose stimuli based on the numerosity rule they had been trained) but had chosen one of these two stimuli significantly more than the other, this would suggest bees had learned and used edge-length instead of numerosity. We provided pairs of stimuli that varied in this way for edge-length, area, convex hull, spatial frequency and illusionary contour (Area was kept constant throughout training and tests and therefore did not need to be tested). We must keep in mind, as pointed out above, that even when this type of design suggests continuous cues were not used, as it had in our work, other strategies could still be used. Although bees' behaviour [79] indicated some form of counting, the bumblebees could have used working spatial memory to avoid recently visited shapes (cf. "inhibition of return" [80,81]). Therefore, it is possible that bees discriminated stimuli based on duration of time taken to scan all shapes within a display, or perhaps by an accumulator mechanism responding to visual changes as they scanned past each shape [69]. Either of these possible strategies do not require a true sense of number.

(c) Ways forward

How then can we address this natural, deep-seated issue? We propose that the method of assessing all continuous cues in one unrewarded test, in conjunction with varying all continuous cues during training, be set as a minimum when investigating numerical cognition in animals. But, as mentioned above we need to still keep in mind other potential non-numerical strategies.

Most numerical cognition studies utilise visual stimuli. Stimuli in other modalities come with their own set of issues regarding continuous variables. For example, number of individuals covaries with the overall complexity of their chemical/olfactory cues, and with the total volume and complexity of vocal calls. However, combining modalities does offer some promising avenues for investigation. One of the strongest pieces of evidence for numerical cognition is the ability to transfer across modalities, which seems to prevent the use of continuous cues because the only similarity across modalities should be numerosity. A nice example of this was shown in monkeys where they were able to match the sum of randomly-ordered sequentially-presented shapes and tones to a visual array with the same number of squares [82]. This kind of cross-modal generalization design would certainly strengthen arguments for numerical cognition in other animals.

Video of animals solving numerical cognition tasks can help determine how animals are solving those tasks (cf. [1,2]). Automated approaches combining machine vision and learning with computational behavioural analyses have the ability to discover behavioural features that humans cannot (cf. [83]; [84]). For example, by measuring the inspection behaviour (e.g. gaze, body direction, movement) of an animal towards different numerical stimuli and comparing across different decisions (choose/reject) and different outcomes (correct/incorrect), underlying strategies may become apparent.

Ultimately, however, we must establish the underlying neural mechanisms to truly know which cues and strategies an animal utilised to solve a numeric-based task. This will provide vital information for how numerical cognition may have evolved, and how processing of numerosity compares between animals [85,86].

Author contributions

HM, ABB and CS conceptualised the project and designed experiments. SL, MH, OJL and CS conducted experiments with the help of FP and WL. CS analysed the behavioural data with help from FP. HM analysed visual stimuli and created and analysed the model with helpful comments from ABB, JARM, AC and EV. HM, ABB and CS wrote the paper with helpful comments from the other authors.

Acknowledgements

This study was supported by the EPSRC program grant Brains-on-Board (EP/P006094/1) awarded to JARM and EV. ABB and JARM were supported by a Leverhulme visiting professorship. ABB and CS were supported by the Templeton World Charity Foundation project number TWCF0266. FP was supported by the National Natural Science Foundation of China (Project no. 31700988). We thank Yonghe Zhou and Yuyi Lu for assistance with the experimental setup in China.

Conflict of interest statement

All authors declare no conflict of interest.

Data availability

The data supporting the findings of this study (figure 1*f*, *g*, *h* and *j*, figure 2*b*, and figure 3*b*, *c*, *d*, and *e*), the code necessary for the model, and the code for measuring the continuous visual features of the stimuli are available in the public repository figshare at https://figshare.com/s/21c5753e31f51ece5f1c. Please note that this is a private link for referees and will be made public upon publication.

References

1. MaBouDi H, Solvi C, Chittka L. 2020 Bumblebees Learn a Relational Rule but Switch to a Win-Stay/Lose-Switch Heuristic After Extensive Training. *Front. Behav. Neurosci.* **14**. (doi:10.3389/fnbeh.2020.00137)

2. Guiraud M, Roper M, Chittka L. 2018 High-speed videography reveals how honeybees can turn a spatial concept learning task into a simple discrimination task by stereotyped

flight movements and sequential inspection of pattern elements. *Front. Psychol.* **9**, 1347. (doi:10.3389/fpsyg.2018.01347)

3. Izquierdo A, Belcher AM. 2012 Rodent Models of Adaptive Decision Making. *Methods Mol. Biol. Clifton NJ* **829**, 85–101. (doi:10.1007/978-1-61779-458-2_5)

4. Risko EF, Gilbert SJ. 2016 Cognitive Offloading. *Trends Cogn. Sci.* **20**, 676–688. (doi:10.1016/j.tics.2016.07.002)

5. Jolicoeur P. 1988 Mental rotation and the identification of disoriented objects. *Can. J. Psychol.* **42**, 461–478. (doi:10.1037/h0084200)

6. Wasserman EA, Zentall TR. 2006 *Comparative Cognition: Experimental Explorations of Animal Intelligence*. Oxford University Press.

7. Chittka L, Rossiter SJ, Skorupski P, Fernando C. 2012 What is comparable in comparative cognition? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 2677–2685. (doi:10.1098/rstb.2012.0215)

8. Cammaerts M-C, Cammaerts R. 2019 Ants Correctly Locate the Zero in a Continuous Series of Numbers. *Int. J. Biol.* **11**. (doi:10.5539/ijb.v11n4p16)

9. Vonk J, Beran MJ. 2012 Bears ‘count’ too: quantity estimation and comparison in black bears, *Ursus americanus*. *Anim. Behav.* **84**, 231–238. (doi:10.1016/j.anbehav.2012.05.001)

10. Aulet LS, Chiu VC, Prichard A, Spivak M, Lourenco SF, Berns GS. 2019 Canine sense of quantity: evidence for numerical ratio-dependent activation in parietotemporal cortex. *Biol. Lett.* **15**, 20190666. (doi:10.1098/rsbl.2019.0666)

11. Rugani R, Vallortigara G, Priftis K, Regolin L. 2015 Number-space mapping in the newborn chick resembles humans’ mental number line. *Science* **347**, 534–536. (doi:10.1126/science.aaa1379)

12. Rugani R, Vallortigara G, Regolin L. 2013 Numerical abstraction in young domestic chicks (*Gallus gallus*). *PLOS ONE* **8**, e65262. (doi:10.1371/journal.pone.0065262)

13. Rugani R, McCrink K, de Hevia M-D, Vallortigara G, Regolin L. 2016 Ratio abstraction over discrete magnitudes by newly hatched domestic chicks (*Gallus gallus*). *Sci. Rep.* **6**, 1–8. (doi:10.1038/srep30114)

14. Tomonaga M. 2008 Relative numerosity discrimination by chimpanzees (*Pan troglodytes*): evidence for approximate numerical representations. *Anim. Cogn.* **11**, 43–57. (doi:10.1007/s10071-007-0089-0)

15. Ditz HM, Nieder A. 2016 Numerosity representations in crows obey the Weber–Fechner law. *Proc. R. Soc. B Biol. Sci.* **283**, 20160083. (doi:10.1098/rspb.2016.0083)

16. Wagener L, Loconsole M, Ditz HM, Nieder A. 2018 Neurons in the Endbrain of Numerically Naive Crows Spontaneously Encode Visual Numerosity. *Curr. Biol.* **28**, 1090-1094.e4. (doi:10.1016/j.cub.2018.02.023)

17. Ditz HM, Nieder A. 2015 Neurons selective to the number of visual items in the corvid songbird endbrain. *Proc. Natl. Acad. Sci.* **12**, 7827–7832. (doi:10.1073/pnas.1504245112)

18. Yaman S, Kilian A, von Fersen L, Güntürkün O. 2012 Evidence for a Numerosity Category that is Based on Abstract Qualities of “Few” vs. “Many” in the Bottlenose Dolphin (*Tursiops truncatus*). *Front. Psychol.* **3**. (doi:10.3389/fpsyg.2012.00473)

19. Irie N, Hiraiwa-Hasegawa M, Kutsukake N. 2019 Unique numerical competence of Asian elephants on the relative numerosity judgment task. *J. Ethol.* **37**, 111–115. (doi:10.1007/s10164-018-0563-y)

20. Piffer L, Petrazzini MEM, Agrillo C. 2013 Large number discrimination in newborn fish. *PLOS ONE* **8**, e62466. (doi:10.1371/journal.pone.0062466)

21. Agrillo C, Miletto Petrazzini ME, Tagliapietra C, Bisazza A. 2012 Inter-Specific Differences in Numerical Abilities Among Teleost Fish. *Front. Psychol.* , 483. (doi:10.3389/fpsyg.2012.00483)

22. DeLong CM, Barbato S, O’Leary T, Wilcox KT. 2017 Small and large number discrimination in goldfish (*Carassius auratus*) with extensive training. *Behav. Processes* **141**, 172–183. (doi:10.1016/j.beproc.2016.11.011)

23. Vonk J, Torgerson-White L, McGuire M, Thueme M, Thomas J, Beran MJ. 2014 Quantity estimation and comparison in western lowland gorillas (*Gorilla gorilla gorilla*). *Anim. Cogn.* **17**, 755–765. (doi:10.1007/s10071-013-0707-y)

24. Lucon-Xiccato T, Miletto Petrazzini ME, Agrillo C, Bisazza A. 2015 Guppies discriminate between two quantities of food items but prioritize item size over total amount. *Anim. Behav.* **107**, 183–191. (doi:10.1016/j.anbehav.2015.06.019)

25. Bisazza A, Agrillo C, Lucon-Xiccato T. 2014 Extensive training extends numerical abilities of guppies. *Anim. Cogn.* **17**, 1413–1419. (doi:10.1007/s10071-014-0759-7)

26. Gatto E, Lucon-Xiccato T, Savaşçı BB, Dadda M, Bisazza A. 2017 Experimental setting affects the performance of guppies in a numerical discrimination task. *Anim. Cogn.* **20**, 187–198. (doi:10.1007/s10071-016-1037-7)

27. Howard SR, Avarguès-Weber A, Garcia JE, Greentree AD, Dyer AG. 2019 Numerical cognition in honeybees enables addition and subtraction. *Sci. Adv.* **5**, eaav0961. (doi:10.1126/sciadv.aav0961)

28. Howard SR, Avarguès-Weber A, Garcia JE, Greentree AD, Dyer AG. 2018 Numerical ordering of zero in honey bees. *Science* **360**, 1124–1126. (doi:10.1126/science.aar4975)

29. Howard SR, Avarguès-Weber A, Garcia JE, Greentree AD, Dyer AG. 2019 Symbolic representation of numerosity by honeybees (*Apis mellifera*): matching characters to small quantities. *Proc. R. Soc. B Biol. Sci.* **286**, 20190238. (doi:10.1098/rspb.2019.0238)

30. Howard SR, Avarguès-Weber A, Garcia JE, Greentree AD, Dyer AG. 2019 Surpassing the subitizing threshold: appetitive–aversive conditioning improves discrimination of numerosities in honeybees. *J. Exp. Biol.* **222**. (doi:10.1242/jeb.205658)

31. Bortot M, Agrillo C, Avarguès-Weber A, Bisazza A, Miletto Petrazzini ME, Giurfa M. 2019 Honeybees use absolute rather than relative numerosity in number discrimination. *Biol. Lett.* **15**, 20190138. (doi:10.1098/rsbl.2019.0138)

32. Howard SR, Schramme J, Garcia JE, Ng L, Avarguès-Weber A, Greentree AD, Dyer AG. 2020 Spontaneous quantity discrimination of artificial flowers by foraging honeybees. *J. Exp. Biol.* **223**. (doi:10.1242/jeb.223610)

33. de Hevia MD, Addabbo M, Nava E, Croci E, Girelli L, Macchi Cassia V. 2017 Infants' detection of increasing numerical order comes before detection of decreasing number. *Cognition* **158**, 177–188. (doi:10.1016/j.cognition.2016.10.022)

34. Jones SM, Pearson J, DeWind NK, Paulsen D, Tenekedjieva A-M, Brannon EM. 2014 Lemurs and macaques show similar numerical sensitivity. *Anim. Cogn.* **17**, 503–515. (doi:10.1007/s10071-013-0682-3)

35. Okuyama S, Iwata J, Tanji J, Mushiake H. 2013 Goal-oriented, flexible use of numerical operations by monkeys. *Anim. Cogn.* **16**, 509–518. (doi:10.1007/s10071-012-0592-9)

36. Ramirez-Cardenas A, Moskaleva M, Nieder A. 2016 Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network. *Curr. Biol.* **26**, 1285–1294. (doi:10.1016/j.cub.2016.03.052)

37. Reznikova Z, Panteleeva S, Vorobyeva NS. 2019 Precise relative-quantity judgement in the striped field mouse Apodemus agrarius Pallas. *Anim. Cogn.* **22**, 277–289. (doi:10.1007/s10071-019-01244-7)

38. Scarf D, Hayne H, Colombo M. 2011 Pigeons on Par with Primates in Numerical Competence. *Science* **334**, 1664–1664. (doi:10.1126/science.1213357)

39. Lucon-Xiccato T, Gatto E, Bisazza A. 2018 Quantity discrimination by treefrogs. *Anim. Behav.* **139**, 61–69. (doi:10.1016/j.anbehav.2018.03.005)

40. Vallortigara G. 2017 An animal's sense of number. In *The nature and development of mathematics: cross disciplinary perspectives on cognition, learning and culture.*, pp. 43–65.

41. Bortot M, Regolin L, Vallortigara G. 2020 A sense of number in invertebrates. *Biochem. Biophys. Res. Commun.* (doi:10.1016/j.bbrc.2020.11.039)

42. Nieder A, Dehaene S. 2009 Representation of Number in the Brain. *Annu. Rev. Neurosci.* **32**, 185–208. (doi:10.1146/annurev.neuro.051508.135550)

43. Henik A. 2016 *Continuous issues in numerical cognition: how many or how much*. Academic Press.

44. 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. *Behav. Brain Sci.* **40**, e164. (doi:10.1017/S0140525X16000960)

45. Banks MS, Ginsburg AP. 1985 Infant visual preferences: A review and new theoretical treatment. In *Advances in child development and behavior* (ed HW Reese), pp. 207–246. JAI. (doi:10.1016/S0065-2407(08)60392-4)

46. Mix KS, Huttenlocher J, Levine SC. 2002 Multiple cues for quantification in infancy: is number one of them? *Psychol. Bull.* **128**, 278–294. (doi:10.1037/0033-2909.128.2.278)

47. Maffei L, Fiorentini A. 1977 Spatial frequency rows in the striate visual cortex. *Vision Res.* **17**, 257–264. (doi:10.1016/0042-6989(77)90089-X)

48. Pollen DA, Ronner SF. 1983 Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. Syst. Man Cybern.* **SMC-13**, 907–916. (doi:10.1109/TSMC.1983.6313086)

49. Horridge GA. 1997 Pattern discrimination by the honeybee: disruption as a cue. *J. Comp. Physiol. A* **181**, 267–277. (doi:10.1007/s003590050113)

50. Lehrer M. 1999 Shape Perception in the Honeybee: Symmetry as a Global Framework. *Int. J. Plant Sci.* **160**, S51–S65. (doi:10.1086/314216)

51. Lehrer M. 1997 Honeybee’s Use of Spatial Parameters for Flower Discrimination. *Isr. J. Plant Sci.* **45**, 157–167. (doi:10.1080/07929978.1997.10676681)

52. Dafni A, Lehrer M, Kevan PG. 1997 Spatial flower parameters and insect spatial vision. *Biol. Rev.* **72**, 239–282. (doi:10.1017/S0006323196005002)

53. Rivas-Blanco D, Pohl I-M, Dale R, Heberlein MTE, Range F. 2020 Wolves and Dogs May Rely on Non-numerical Cues in Quantity Discrimination Tasks When Given the Choice. *Front. Psychol.* **11**. (doi:10.3389/fpsyg.2020.573317)

54. Miletto Petrazzini ME, Wynne CDL. 2016 What counts for dogs (Canis lupus familiaris) in a quantity discrimination task? *Behav. Processes* **122**, 90–97. (doi:10.1016/j.beproc.2015.11.013)

55. Gatto E, Carlesso D. 2019 Spontaneous quantity discrimination in crickets. *Ethology* **125**, 613–619. (doi:10.1111/eth.12912)

56. Leibovich T, Henik A. 2014 Comparing performance in discrete and continuous comparison tasks. *Q. J. Exp. Psychol.* **67**, 899–917. (doi:10.1080/17470218.2013.837940)

57. Rugani R, Regolin L, Vallortigara G. 2010 Imprinted numbers: newborn chicks’ sensitivity to number vs. continuous extent of objects they have been reared with. *Dev. Sci.* **13**, 790–797. (doi:10.1111/j.1467-7687.2009.00936.x)

58. Agrillo C, Piffer L, Bisazza A. 2011 Number versus continuous quantity in numerosity judgments by fish. *Cognition* **119**, 281–287. (doi:10.1016/j.cognition.2010.10.022)

59. Burnham KP, Anderson DR. 2002 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edn. New York: Springer-Verlag. (doi:10.1007/b97636)

60. van der Schaaf A, van Hateren JH. 1996 Modelling the Power Spectra of Natural Images: Statistics and Information. *Vision Res.* **36**, 2759–2770. (doi:10.1016/0042-6989(96)00002-8)

61. O'Carroll D. 1993 Feature-detecting neurons in dragonflies. *Nature* **362**, 541–543. (doi:10.1038/362541a0)

62. James AC, Osorio D. 1996 Characterisation of columnar neurons and visual signal processing in the medulla of the locust optic lobe by system identification techniques. *J. Comp. Physiol. A* **178**, 183–199. (doi:10.1007/BF00188161)

63. Sachs MB, Nachmias J, Robson JG. 1971 Spatial-Frequency Channels in Human Vision*. *JOSA* **61**, 1176–1186. (doi:10.1364/JOSA.61.001176)

64. Silverman MS, Grosof DH, De Valois RL, Elfar SD. 1989 Spatial-frequency organization in primate striate cortex. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 711–715. (doi:10.1073/pnas.86.2.711)

65. Srinivasan MV. 2010 Honey Bees as a Model for Vision, Perception, and Cognition. *Annu. Rev. Entomol.* **55**, 267–284. (doi:10.1146/annurev.ento.010908.164537)

66. Gevers W, Kadosh RC, Gebuis T. 2016 Sensory integration theory: an alternative to the approximate number system. In *Continuous issues in numerical cognition: how many or how much* (ed A Henik), pp. 405–418. San Diego: Academic Press.

67. Uller C, Lewis J. 2009 Horses (*Equus caballus*) select the greater of two quantities in small numerical contrasts. *Anim. Cogn.* **12**, 733–738. (doi:10.1007/s10071-009-0225-0)

68. Ditz HM, Nieder A. 2020 Format-dependent and format-independent representation of sequential and simultaneous numerosity in the crow endbrain. *Nat. Commun.* **11**, 1–10. (doi:10.1038/s41467-020-14519-2)

69. Meck WH, Church RM. 1983 A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 320–334. (doi:10.1037/0097-7403.9.3.320)

70. Gallistel CR, Gelman R. 2000 Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci.* **4**, 59–65. (doi:10.1016/s1364-6613(99)01424-2)

71. Vasas V, Chittka L. 2019 Insect-Inspired Sequential Inspection Strategy Enables an Artificial Network of Four Neurons to Estimate Numerosity. *iScience* **11**, 85–92. (doi:10.1016/j.isci.2018.12.009)

72. Rapp H, Nawrot MP, Stern M. 2020 Numerical Cognition Based on Precise Counting with a Single Spiking Neuron. *iScience* **23**, 100852. (doi:10.1016/j.isci.2020.100852)

73. Eimas PD. 1969 Multiple-Cue Discrimination Learning in Children. *Psychol. Rec.* **19**, 417–424. (doi:10.1007/BF03393868)

74. Slotnick B. 2001 Animal cognition and the rat olfactory system. *Trends Cogn. Sci.* **5**, 216–222. (doi:10.1016/S1364-6613(00)01625-9)

75. Gerber B, Smith BH. 1998 Visual modulation of olfactory learning in honeybees. *J. Exp. Biol.* **201**, 2213–2217.

76. Ohnishi K. 1991 Goldfish's visual information processing patterns in food-reinforced discrimination learning between compound visual stimuli. *J. Comp. Physiol. A* **168**, 581–589. (doi:10.1007/BF00215080)

77. Brembs B, Heisenberg M. 2001 Conditioning with compound stimuli in drosophila melanogaster in the flight simulator. *J. Exp. Biol.* **204**, 2849–2859.

78. Clipperton-Allen A, Cole M, Peck M, Quirt J. 2016 Pattern cue and visual cue competition in a foraging task by rats. *Learn. Behav.* **44**, 378–389. (doi:10.3758/s13420-016-0231-4)

79. MaBouDi H, Dona HSG, Gatto E, Loukola OJ, Buckley E, Onoufriou PD, Skorupski P, Chittka L. 2020 Bumblebees use sequential scanning of countable items in visual patterns to solve numerosity tasks. *Integr. Comp. Biol.* (doi:10.1093/icb/icaa025)

80. Gabay S, Leibovich T, Ben-Simon A, Henik A, Segev R. 2013 Inhibition of return in the archer fish. *Nat. Commun.* **4**, 1657. (doi:10.1038/ncomms2644)

81. Sapir A, Soroker N, Berger A, Henik A. 1999 Inhibition of return in spatial attention: direct evidence for collicular generation. *Nat. Neurosci.* **2**, 1053–1054. (doi:10.1038/15977)

82. Jordan KE, MacLean EL, Brannon EM. 2008 Monkeys match and tally quantities across senses. *Cognition* **108**, 617–625. (doi:10.1016/j.cognition.2008.05.006)

83. Egnor SER, Branson K. 2016 Computational Analysis of Behavior. *Annu. Rev. Neurosci.* **39**, 217–236. (doi:10.1146/annurev-neuro-070815-013845)

84. Skorupski P, MaBouDi H, Dona HSG, Chittka L. 2018 Counting insects. *Phil Trans R Soc B* **373**, 20160513. (doi:10.1098/rstb.2016.0513)

85. Nieder A. 2016 The neuronal code for number. *Nat. Rev. Neurosci.* **17**, 366–382. (doi:10.1038/nrn.2016.40)

86. Harvey BM. 2016 Quantity Cognition: Numbers, Numerosity, Zero and Mathematics. *Curr. Biol.* **26**, R419–R421. (doi:10.1016/j.cub.2016.03.059)

Figures and legends

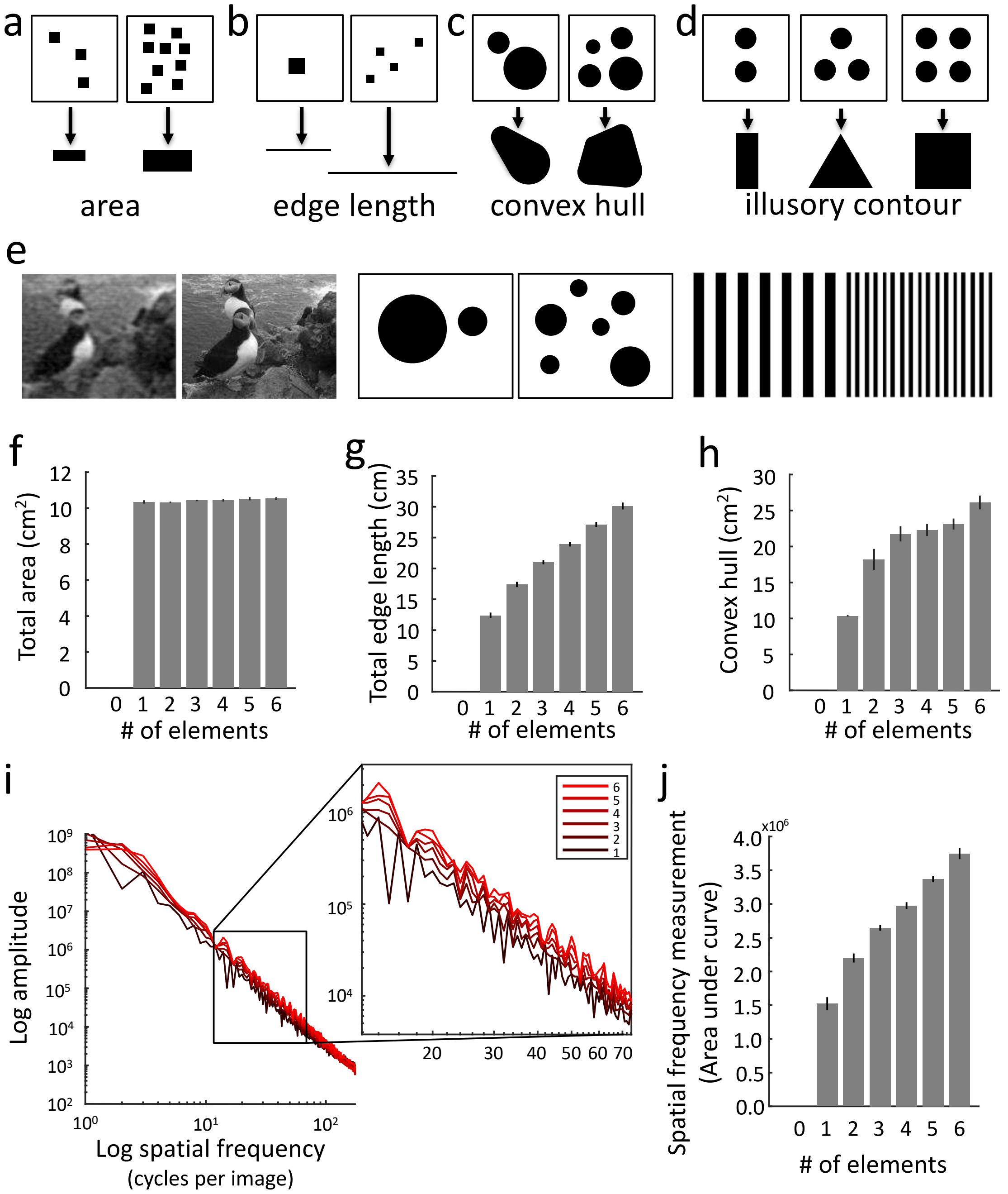

Figure 1. Number of elements naturally co-varies with non-numerical cues. (a-d) Examples of 2D stimuli used in numerical cognition studies and how different continuous cues normally covary with numerosity. Note that illusory contour does not covary with numerosity but can still be learned and used to solve numerical cognition tasks, especially with lower number of elements. **(e)** Spatial frequency (the amount of alternating dark and light regions in a given area) also normally covaries with numerosity. The more changes from black to white across an image in all directions, the greater spatial frequency. The right images of each pair in (e) all have higher spatial frequency than the left images. **(f-h)** For all stimuli in [28], from which our stimulus set was borrowed, area (amount of total black (inset)) was kept constant (f), but edge length (total boundary length (inset); g) and convex hull (the minimum convex region covering all elements (inset); h) covaried with numerosity. **(j-i)** Spatial

frequency is calculated by obtaining a power spectrum (Methods) and measuring the area under the power spectrum's curve. The power spectrum plots (f and zoomed-in inset) for all stimuli in [28], from which our stimulus set was borrowed, averaged for each number of elements from one to six, shows that spatial frequency increases with numerosity (g). Note that for all covarying continuous cues, a zero-set stimulus will have zero measurement and thereby be placed naturally at the lower end of the spectrum for each of these non-numerical cues.

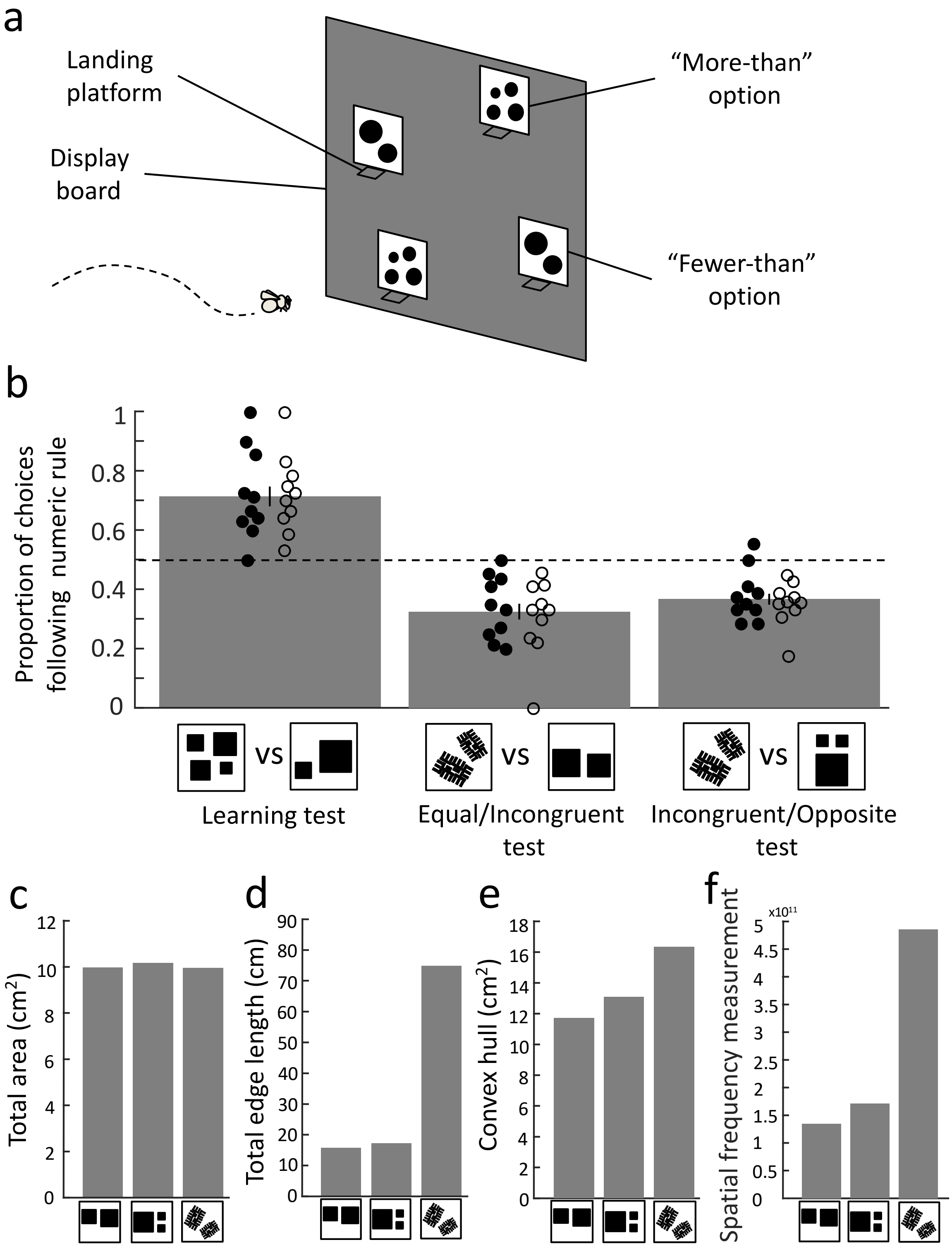

Figure 2. Bees can use non-numerical strategies to discriminate numerical stimuli. **(a)** Experiment setup. Honeybees were trained to find 50% sucrose solution at one of two pairs of displays showing different numbers of elements, and aversive quinine solution on the other display pair (Methods). **(b)** Once honeybees reached 80% performance, they were tested using displays with novel shapes. In the Learning test, honeybees more often chose stimuli following the numerical rule on which they had been trained (71.3±3.3%; more-than: 70.3±4.7%; less-than: 72.4±4.8%). However, when tested on stimuli that differed in continuous cues but not

number of elements (Equal/Incongruent test; middle bar; 32.5±2.6; more-than: 30.7±4.2%; less-than: 34.2±3.4%) and separately on two pairs of stimuli where numerosity and continuous cues were set in opposition (Incongruent/Opposite test; right bar; 36.7±1.8; more-than: 35.1±2.4%; less-than: 38.2±2.8%), honeybees chose stimuli based on continuous cues over numerosity. Data shown are combined from the two groups trained with different numerical rules since no difference in performance was found between groups (Table 1; Methods). Dotted line = 0.5 chance level. Bars = mean. Vertical lines = s.e.m. Circles = individual bees' data points (filled circles: bees trained to more-than rule; empty circles: bees trained to less-than rule). **(c-f)** Stimuli used in tests with corresponding continuous variable measurements (Methods).

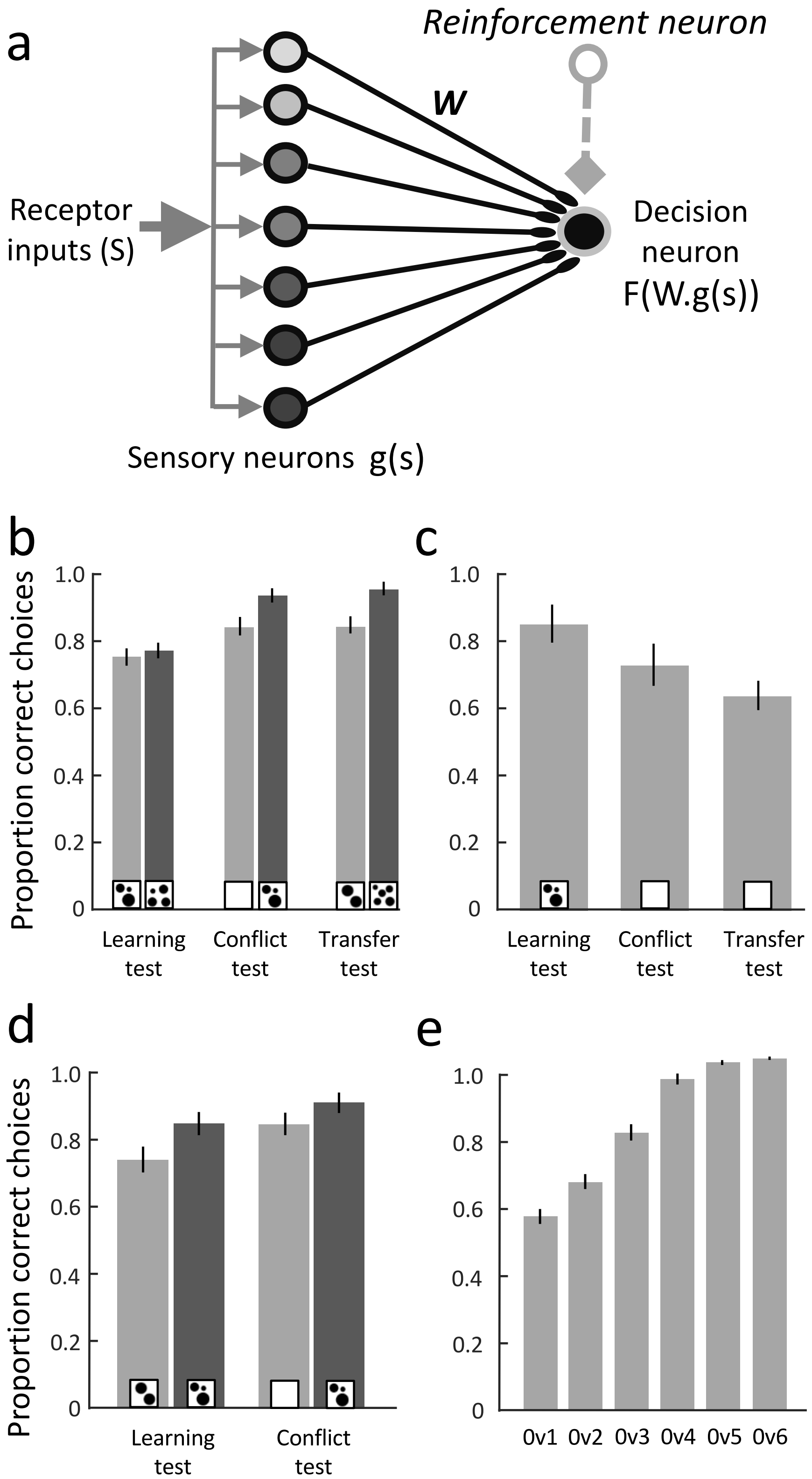

Figure 3. A simple computational model using only non-numerical cues reproduces honeybees' performance on a numerosity task. (a) The model uses seven sensory neurons that are activated by the output of visual receptors. Each sensory neuron responds to multiple levels of a single continuous cue with different sensitivities. Firing of each sensory neuron is specific and selective to the preference

level modelled by a Gaussian tuning curve. Information from all sensory neurons converges at a single decision neuron. Synaptic connectivity (W) between sensory neurons and the decision neuron are modified by an associative learning rule for encoding appetitive and aversive valences. Performance of the model is evaluated by a simple subtraction of the responses of the decision neuron to the test stimuli. Our model is able to reproduce behaviours claimed to be indicative of numerical cognition [28], without any reference to numerosity. This includes transferring a "more-than" or "less-than" rule to: **(b)** novel shapes in a "conflict test" examining preference for zero numerosity (Wilcoxon signed rank test, z-value>6.22 and p<3.50e-9) and a "transfer test" using displays with more shapes than in training (Wilcoxon signed rank test, z-value>7.99 and p<3.17e-14). Compare to [28] Fig 1C. **(c)** stimuli containing a number of elements outside the training stimuli range, in a learning test (Wilcoxon signed rank test, z-value=3.89 and p=9.98e-05), conflict test (z-value=3.23 and p=0.0012) and transfer test (z-value=2.40 and p=0.016). Compare to [28] Figure 1D). **(d)** novel pairs of stimuli, including stimuli with zero elements, in a learning test (Wilcoxon signed rank test, z-values> 5.27 and p<1.35e-06), and conflict test (Wilcoxon signed rank test, z-values> 5.51 and p<3.49e-07). Compare to [28] Figure S4. **(E)** and recognising stimuli with zero elements as the lower end of a continuum (Wilcoxon signed rank test for comparing each pair with the chance level 50%, z-values> 2.24 and p<0.024; Kruskal-Wallis test, df=299; chi-sq=183.94 and p=7.71e-37. Compare to [28] Figure 2B. Light grey=less-than. Dark grey=more-than. Insets=test stimuli. Bars=mean. Vertical lines=s.e.m. calculated from the firing rate of the decision neuron for 50 different initial parameters that simulated 50 different model bees.

Table S1. Summary of the full generalized linear mixed models examining factors in relation to proportion of correct choices during each test

Dependent variable	Fixed factors	df	Estimate	SE	*z-value*	*P*	*AICc*
Proportion correct in Learning test	Intercept	16	0.66	0.61	1.09	0.28	76.7
	Rule	16	0.19	0.29	0.65	0.52	
	Country	16	-0.11	0.30	-0.38	0.70	
Proportion correct in Equal/Incongruent test	Intercept	16	-0.68	0.57	-1.19	0.23	80.7
	Rule	16	0.91	0.26	0.35	0.72	
	Country	16	-0.06	0.26	-0.22	0.83	
Proportion correct in Incongruent/Opposite test	Intercept	16	-0.69	0.57	-1.22	0.22	76.6
	Rule	16	0.11	0.25	0.43	0.67	
	Country	16	-0.01	0.26	-0.04	0.97	

For each glmm, the dependent variable was the proportion of correct choices during that test (following the numeric rule on which the bee was trained). The rule (more-than/less-than) and country where the individual bee completed training and tests (China/Finland) were included as fixed factors. Bee ID was used as a random factor.

Table S2. Summary of the reduced generalized linear mixed models examining factors in relation to proportion of correct choices during each test

Dependent variable	Fixed factors	df	Estimate	SE	*z-value*	*P*	*AICc*
Proportion correct in Learning test	Intercept	18	0.66	0.14	5.25	1.49e-7	71.3
Proportion correct in Equal/Incongruent test	Intercept	18	-0.64	0.13	-4.98	6.50e-7	74.9
Proportion correct in Incongruent/Opposite test	Intercept	18	-0.55	0.12	-4.38	1.17e-5	70.8

For each glmm, the dependent variable was the proportion of correct choices during that test (following the numeric rule on which the bee was trained). Bee ID was used as a random factor.