SHORT COMMUNICATION

Surpassing the subitizing threshold: appetitive–aversive conditioning improves discrimination of numerosities in honeybees

Scarlett R. Howard[1,2,*], Aurore Avarguès-Weber[1], Jair E. Garcia[2], Andrew D. Greentree[3] and Adrian G. Dyer[2,4]

ABSTRACT

Animals including humans, fish and honeybees have demonstrated a quantity discrimination threshold at four objects, often known as subitizing elements. Discrimination between numerosities at or above the subitizing range is considered a complex capacity. In the current study, we trained and tested two groups of bees on their ability to differentiate between quantities (4 versus 5 through to 4 versus 8) when trained with different conditioning procedures. Bees trained with appetitive (reward) differential conditioning demonstrated no significant learning of this task, and limited discrimination above the subitizing range. In contrast, bees trained using appetitive–aversive (reward–aversion) differential conditioning demonstrated significant learning and subsequent discrimination of all tested comparisons from 4 versus 5 to 4 versus 8. Our results show conditioning procedure is vital to performance on numerically challenging tasks, and may inform future research on numerical abilities in other animals.

KEY WORDS: Approximate number system, Learning, Number, Numeric, Object file system, Quantity discrimination

INTRODUCTION

Quantity discrimination has been explored in a number of species through different methods (e.g. spontaneous choice versus training; limited versus extensive training), which sometimes yield different results (Agrillo and Bisazza, 2014; DeWind and Brannon, 2012; Gatto et al., 2017; Gazes et al., 2018; Miletto Petrazzini et al., 2018). The ability to discriminate between quantities varies across different species, with insects such as mealworm beetles able to discriminate ratios (calculated by dividing the lower number by the higher number, such as 1 versus 4 and 1 versus 3) of 0.25–0.33 (Carazo et al., 2009). Species that can discriminate the more difficult ratio of 0.80 include monkeys (Addessi et al., 2008; Gazes et al., 2018), jungle crows (Bogale et al., 2011), Shetland ponies (Gabor and Gerken, 2018), Clark's nutcrackers (Tornick et al., 2015) and dolphins (Jaakkola et al., 2005). Some animals including elephants (Irie-Sugimoto et al., 2009), great apes (Hanus and Call, 2007), Mexican jays (Kelly, 2016), western scrub jays (Kelly, 2016) and

guppies (Bisazza et al., 2014; Lucon-Xiccato et al., 2017) can discriminate ratios over 0.80. Previous studies on quantity discrimination suggest some variation between low quantities (4 and fewer) and high quantities (4 and greater). For example, frogs (*Bombina orientalis*) are able to discriminate ratios of 0.67 with 4 items or fewer but only ratios of 0.5 with 4 or more items (Stancher et al., 2015). Discrimination ability can also vary with different number comparisons of the same ratio; for example, North Island robins can discriminate a ratio of 0.875 when comparing quantities of 7 versus 8, but not with quantities of 14 versus 16, despite the two comparisons having the same ratio (Garland et al., 2012).

Several species, including humans, exhibit a numerical competency threshold at 4 (Agrillo et al., 2008; Cowan, 2010; Gómez-Laplaza and Gerlai, 2011a; Jevons, 1871; Kaufman et al., 1949; Simons and Langheinrich, 1982; Trick and Pylyshyn, 1994). This threshold is observed in visual object and tactile stimulation estimation in humans (Dehaene and Cohen, 1994; Jevons, 1871; Piazza et al., 2002; Riggs et al., 2006; Starkey and Cooper, 1995), quantity discrimination in fish (Agrillo et al., 2008; Gómez-Laplaza and Gerlai, 2011a; Seguin and Gerlai, 2017), as well as proto-counting and number-generalization in honeybees (Boysen, 1988; Dacke and Srinivasan, 2008; Gross et al., 2009). The threshold phenomenon is observed across a diverse range of species and may be due to the hypothesis that quantity/numerical discrimination has two mechanisms: 'subitizing'/object file system (accurately and quickly recognizing 4 or fewer objects) and counting (the process of sequentially incrementing the number of identified elements, theoretically without bound; Agrillo et al., 2008; Kaufman et al., 1949; Piazza et al., 2002; Tomonaga and Matsuzawa, 2002; Trick and Pylyshyn, 1994). Some non-human animal species have been shown to surpass this threshold limit of 4 following mechanisms consistent with Weber's law using the approximate number system (ANS). The ANS is a non-symbolic numerical quantification system used by humans and non-human animals for quantifying numerosities above 4 (Agrillo et al., 2008; DeWind et al., 2015; Feigenson, 2004; Vallortigara, 2017). Weber's law describes how sensory systems distinguish between two stimulus magnitudes based on their proportional difference (Akre and Johnsen, 2014; Fechner, 1965; Weber, 1978). For example, the ability of mosquitofish to discriminate between two groups of conspecifics consisting of more than 4 individuals improves as the numerical distance increases; individuals fail at 4 versus 5 but succeed at 4 versus 8 (Agrillo et al., 2008).

While many studies have explored the numerical abilities of vertebrates, relatively little research has been done on the numerical competency of invertebrates (Carazo et al., 2009; Dacke and Srinivasan, 2008). Studies testing the ability of bees to 'proto-count' landmarks in either open natural environments or controlled laboratory conditions found that bees were able to reliably

[1]Centre de Recherches sur la Cognition Animale, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, 31400 Toulouse, France. [2]Bio-inspired Digital Sensing (BIDS) Lab, School of Media and Communication, RMIT University, Melbourne, VIC 3000, Australia. [3]ARC Centre of Excellence for Nanoscale BioPhotonics, School of Science, RMIT University, Melbourne, VIC 3000, Australia. [4]Department of Physiology, Monash University, Clayton, VIC 3800, Australia.

*Author for correspondence (scarlett.howard@rmit.edu.au)

S.R.H., 0000-0002-1895-5409

numerically assess up to 4 landmarks. True counting requires the abstraction principle to be fulfilled, where counting should be able to be transferred between different object types (Chittka and Geiger, 1995); thus, counting in honeybees is termed proto-counting. Bees failed at proto-counting more than 4 landmarks in order to find a source of food (Chittka and Geiger, 1995; Dacke and Srinivasan, 2008). In a study using a delayed matching to sample (DMTS) procedure in a Y-maze, bees were able to accurately match specific quantities of up to 3 elements irrespective of shape, colour and pattern of the objects to receive a reward of sucrose, but were unable to do this with higher numbers (Gross et al., 2009). Bees showed a discrimination ability of 2 versus 3 and 3 versus 4 but not 4 versus 5 or 4 versus 6. These studies were important steps in determining that bees could both proto-count and match numbers, and that numerical ability was a biologically meaningful skill for either navigation (Chittka and Geiger, 1995; Dacke and Srinivasan, 2008) or numerically assessing the number of flowers in a flower patch during foraging (Gross et al., 2009).

The previous studies on number matching and proto-counting in honeybees used classic appetitive or appetitive differential conditioning frameworks, where bees received a reward of sucrose for a correct choice and no outcome for an incorrect choice (Chittka and Geiger, 1995; Dacke and Srinivasan, 2008; Gross et al., 2009). Advancements in training protocol have shown that when bees are trained on a perceptually difficult colour task, performance significantly increases when appetitive–aversive differential conditioning is used in training compared with appetitive differential conditioning (Avarguès-Weber et al., 2010; Dyer, 2012). Appetitive–aversive differential conditioning is where a bee receives a reward of sucrose for a correct choice and an aversive outcome of quinine for an incorrect choice, whereas appetitive differential conditioning rewards a bee for a correct choice and gives no outcome for an incorrect choice. There is evidence that including a distractor associated with an aversive outcome also improves the strength of the association between the correct option and the reward in vertebrates (Rescorla and Wagner, 1972), which is potentially linked to attentional resources (Avarguès-Weber et al., 2010; Dyer, 2012; Giurfa, 2004; Rescorla and Wagner, 1972). Recent work on spatial visual processing in honeybees showed that learning and behavioural outcomes can be significantly influenced by the conditioning procedure employed during training (Dyer and Griffiths, 2012; Dyer et al., 2005; Giurfa et al., 1999; Howard et al., 2017; Stach and Giurfa, 2005), and so bees have become an important model species for understanding comparative visual perception and cognition.

Recently, honeybees were shown to learn the rules of 'less than' and 'greater than' and apply these rules to the novel numerosities of 0 and 5 using appetitive–aversive differential conditioning (Howard et al., 2018). Bees demonstrated an ability to learn to discriminate between all combinations of the numerosities 0 to 6 during training in one of the experiments, a discrimination ability that has not previously been shown in honeybees (Howard et al., 2018). We also recently trained bees to learn to add or subtract 1 element from an array of elements consisting of 1 to 5 elements. Bees demonstrated significant success in novel problems and thus demonstrated simple arithmetic abilities when trained with appetitive–aversive differential conditioning (Howard et al., 2019a). Honeybees have also demonstrated the capacity to learn to match signs (abstract symbols) with quantities when trained with appetitive–aversive differential conditioning (Howard et al., 2019b). These results are probably due to the use of appetitive–aversive differential conditioning during training and thus in this study we formally

tested and compared the numerical discrimination ability of bees when trained with appetitive or appetitive–aversive differential conditioning. We determined whether numerical discrimination changes with different conditioning procedures, and thus provide some insights into why different quantity discrimination results are observed across a range of animal taxa.

MATERIALS AND METHODS
Study design
We aimed to determine whether there was a difference in quantity discrimination performance between bees that were trained using appetitive differential conditioning (group 1) compared with bees trained with appetitive–aversive differential conditioning (group 2). To do this, we trained bees to choose 4 elements when presented against the quantities 1 to 10. We then tested bees on their ability to discriminate between quantities of 4 versus 5, 6, 7 and 8.

Study species
We used 22 free-flying honeybees (*Apis mellifera* Linnaeus 1758) foragers for this experiment. All bees were marked with a coloured dot on the thorax to identify individuals. A gravity feeder was set up within 20 m of the experiment to provide 10–30% sucrose which attracted a regular number of bees for use in experiments. Bees were randomly assigned to test groups. Honeybee foragers were collected from a number of different hives and gravity feeders providing sucrose solution (5–30%), which were set up to maintain visiting foragers. The bee hives were maintained at Paul Sabatier University in Toulouse, France, where there were over 25 hives to recruit honeybee foragers from.

Apparatus
Individual honeybees were trained to enter a Y-maze (Fig. 1; as described in Avarguès-Weber et al., 2011). The arms of the Y-maze were 40×20×20 cm (L×W×H). Each bee had to fly through an entrance hole (5 cm in diameter) to enter an initial chamber and then fly through another hole into the decision chamber where it would be presented with two different options in each arm of the chamber, one correct and one incorrect (Fig. 1).

Stimuli were presented on grey backgrounds (20×20 cm) located 5 cm away from the decision lines to ensure each element was above the minimum resolution threshold for free-flying honeybees (Srinivasan and Lehrer, 1988). During training in group 1, a 10 µl drop of 50% sucrose solution (appetitive/rewarding outcome) was paired with a correct choice and a 10 µl drop of water (neutral unreinforced outcome) was paired with the incorrect choice (appetitive differential conditioning). During training in group 2, a 10 µl drop of 50% sucrose solution (appetitive/rewarding outcome) was paired with a correct choice and a 60 mmol l^{-1} quinine solution (aversive outcome) was paired with an incorrect choice (appetitive–aversive differential conditioning). Each stimulus had a transparent landing pole located below it which held the drop of sucrose, quinine or water. Poles were replaced when touched by a bee and cleaned with 20% ethanol then water and dried to exclude olfactory cues. The side of correct and incorrect stimuli was randomly changed between choices. If a bee made an incorrect choice and started to imbibe the quinine, it was allowed to fly to the pole in front of the correct stimulus to collect sucrose to maintain motivation; but only the first choice was recorded for statistical analysis following standard procedures (Avarguès-Weber et al., 2015). Once the bee had finished imbibing the sucrose, it was allowed to fly back to the hive if satiated or make another decision by re-entering the maze from behind an opaque screen. During the non-reinforced tests, a
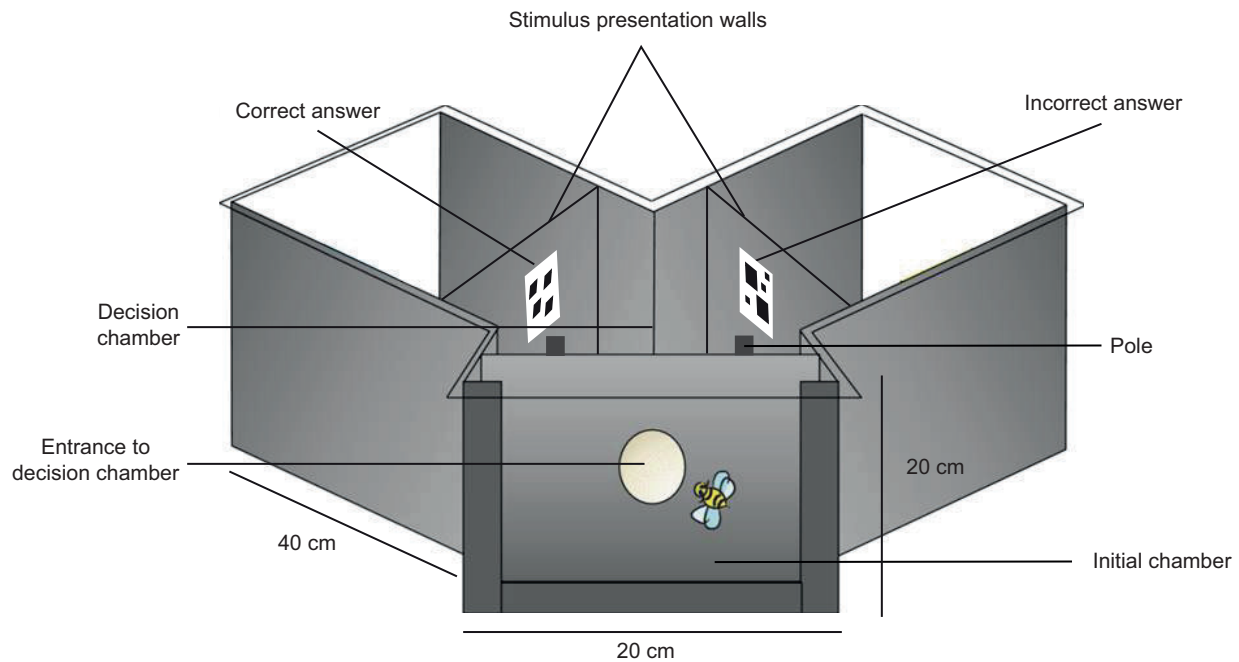
Fig. 1. Y-maze apparatus set-up for the experiments. The diagram shows parts of the Y-maze and the stimulus positions. Bees were trained to choose 4 elements when presented against the incorrect options of 1–10 (excluding 4).

drop of water was placed on each of the poles placed in front of the stimuli. Ten choices (touches of the poles) were recorded for each of the four tests to enable statistical comparisons, consistent with standard testing procedures for honeybees (Avarguès-Weber et al., 2015; Howard et al., 2018, 2017).

Stimuli

Each stimulus was a 6×6 cm white square card containing a number of black elements (Fig. 2) and was covered with 80 µm Lowell laminate. Elements in the learning phase could be one of five shapes: square, diamond, circle, triangle or star. Randomized shapes were used for the four tests to ensure patterns and shapes were unfamiliar to bees during tests (Fig. 3A). Stimuli ranged from 1 to 10 elements, and no stimulus was shown in more than a single bout (return to hive to off-load sucrose; approximately 2–5 choices).

There were three sets of stimuli consisting of (i) equal overall surface area (set 1; n=164; surface area 10±1 cm²), (ii) consistently equal element area (set 2; n=160; all elements were 1 cm²) or (iii) novel randomized shapes (set 3; n=20; Fig. 2). Set 1 had an average element area of 2.039 cm². The training sets (set 1 and set 2) had an overall average surface area of 1.520 cm². There were 344 stimuli in total; furthermore, stimuli were rotated to one of four orientation positions determined randomly by dice roles to provide training and testing sets with over 1000 options. All stimuli were created in Adobe Illustrator and stimuli properties (e.g. set 1 of equal overall surface area and set 2 of equal element area) were validated with the program ImageJ.

The average inter-item distance of stimuli, a measure of density, for each set is presented in Fig. S1. Another measure of density/occupancy (Bertamini et al., 2018) – the mean proportion of black pixels – of stimuli is presented in Fig. S2.

Training procedure

Bees were incrementally trained to enter the Y-maze and both arms of the apparatus over 30–60 min periods. Once each bee was able to fly into the entrance hole and the hole that led to the decision

chamber, and could find the poles in both Y-maze arms, the experiment began.

Each bee completed either 50 appetitive differential conditioning choices (group 1) or 50 appetitive–aversive differential conditioning choices (group 2). Bees in group 1 were rewarded with sucrose for a choice of 4 elements and received no outcome (a drop of water on pole) for an incorrect choice of any other number. Bees in group 2 were rewarded with sucrose for a choice of 4 elements and received an aversive outcome, quinine, for a choice of any other number of elements ranging from 1 to 10 (excluding 4; Fig. 3A).

Testing procedure

Once bees had completed the training, there were four tests of 10 unreinforced choices each for bees in either group. Between each of the four tests there were 10 refresher reinforced choices to maintain bee motivation (same procedure as the learning phase). The sequence of these tests was randomized. The tests were non-reinforced (no reward or punishment) and used a 10 µl drop of water (neutral outcome) instead of quinine or sucrose to motivate bees to land. Bees were shown comparisons of 4 versus 5, 4 versus 6, 4 versus 7 and 4 versus 8 elements (Fig. 3A); these tests compared quantity ratio discriminations of 0.80, 0.67, 0.57 and 0.50, respectively.

Statistical analysis

To test for the effect of training on bee performance (number of correct choices) in both groups, data from the learning phase of 50 choices were analysed with a generalized linear mixed-effect model (GLMM) with a binomial distribution using the 'glmer' package within the R environment for statistical analysis. We fitted a full model with trial number as a continuous predictor, and subject as a random factor to account for repeated choices of individual bees.

To determine whether bees were able to learn to choose 4 elements in tests, we analysed the test data by employing a GLMM including only the intercept term as fixed factor and subject as a
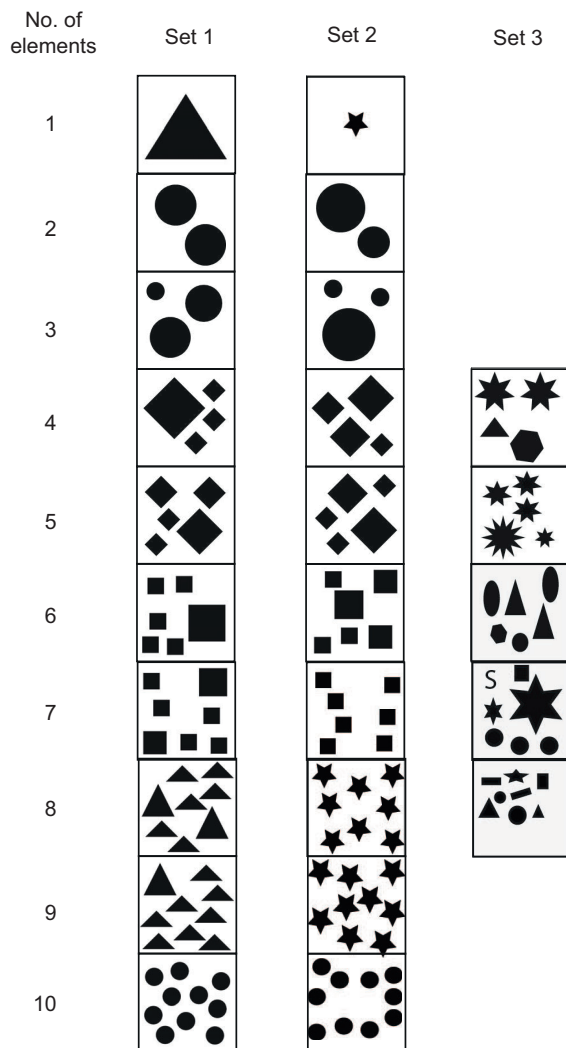
Fig. 2. An example subset of the stimuli used in the training (set 1, set 2) and the testing (set 3) phases. We provide an example of 1 stimulus per quantity per set. Stimuli rotational axis was randomized to one of four positions to further exclude low-level cues.

random term. The proportion of 'correct' choices (MPCC) recorded from the tests was used as the response variable in the model. The Wald statistic (z) tested whether the mean proportion of correct choices recorded from the learning test, represented by the coefficient of the intercept term, was significantly different from chance expectation, i.e. H0: MPCC=0.5. The 95% confidence intervals (CIs) for the model coefficients were constructed using profile likelihood methods rather than calculated from the standard errors of the least-squares estimates (Faraway, 2016).

All analyses were performed within the R environment for statistical analysis.

Ethics

All animal care was in accordance with institutional guidelines.

RESULTS AND DISCUSSION

Two groups of bees were trained for 50 trials of either appetitive differential conditioning (group 1) or appetitive–aversive differential conditioning (group 2; Fig. 3A). Only bees trained using appetitive–aversive differential conditioning demonstrated

significant learning over the period of 50 trials (statistical test: GLMM with a binomial distribution with trial number as a continuous predictor, and subject as a random factor to account for repeated choices of individual bees; group 1: $z=0.102$; $P=0.918$; $n=10$, Fig. 3B; group 2: $z=5.48$; $P<0.001$; $n=12$; Fig. 3C).

After the acquisition phase, honeybees were evaluated on their discrimination ability between 4 elements and higher numerosities in four tests (4 versus 5, 4 versus 6, 4 versus 7 and 4 versus 8), for 10 unreinforced choices per comparison using randomly shaped novel elements (set 3), presented in a random order. Bees trained using an appetitive-only procedure were unable to differentiate between 4 versus 5 (statistical test: logistic regression with individual as random term to test for differences between the observed proportion of bee choices and chance level, $y=0.5$, 51.0±3.8% mean±s.e.m., $z=0.200$, $P=0.841$, 95% CI: 0.413, 0.607), 4 versus 6 (50.0±4.90%, $z=0.000$, $P=1.000$, 95% CI: 0.397, 0.603) or 4 versus 8 (54.0±5.00%, $z=0.600$, $P=0.549$, 95% CI: 0.513, 0.702) at a level significantly different from chance, except during the test of 4 versus 7 (62.0 ±3.60%, $z=2.18$, $P=0.029$, 95% CI: 0.423, 0.637; Fig. 3D). In contrast, bees trained using appetitive–aversive differential conditioning were able to discriminate between 4 versus 5 (59.2±3.10%, $z=2.00$, $P=0.046$, 95% CI: 0.502, 0.677), 4 versus 6 (60.8±3.80%, $z=2.35$, $P=0.019$, 95% CI: 0.519, 0.693), 4 versus 7 (63.3±3.30%, $z=2.89$, $P=0.004$, 95% CI: 0.546, 0.716) and 4 versus 8 (64.2±3.40%, $z=3.06$, $P=0.002$, 95% CI: 0.554, 0.724; Fig. 3E) at a level significantly different from chance expectation.

Our results show that numerical discrimination above the subitizing range is possible with appetitive–aversive differential conditioning. Bees trained using this method were able to discriminate between quantities at a ratio of 0.80 (4 versus 5), a finer discrimination than was previously observed in honeybees considering appetitive conditioning (Gross et al., 2009). There is some evidence that bees trained with appetitive differential conditioning may be able to discriminate ratios of 0.57 (4 versus 7) above the threshold of 4 objects; however, this is still an open question as these bees failed at discriminating the less challenging ratio of 0.50 (4 versus 8). Honeybees are currently at a level of quantity discrimination observed in species such as African grey parrots (Al Aïn et al., 2009), capuchin monkeys (Addessi et al., 2008; Gazes et al., 2018), squirrel monkeys (Gazes et al., 2018), dolphins (Jaakkola et al., 2005), ponies (Gabor and Gerken, 2018), jungle crows (Bogale et al., 2011) and guppies (Bisazza et al., 2014). The question of whether the introduction of an aversive outcome for an incorrect choice improves performance has been asked for colour discrimination tasks in bees (Avarguès-Weber et al., 2010; Chittka et al., 2003), but this is the first time appetitive and appetitive–aversive differential conditioning have been directly compared for a numerical visual task.

A potential explanation for why bees perform better at numerosity discrimination with appetitive–aversive differential conditioning is that the attentional processes are improved as a result of the presence of an aversive outcome for an incorrect choice (Avarguès-Weber et al., 2010). When the penalty for making an incorrect decision is low, there is less motivation to ensure performance is as accurate as possible; however, when the penalty is increased, there is an increase in motivation to be more accurate and thus attention may be modulated by conditioning procedure (Avarguès-Weber et al., 2010). By pairing the incorrect option with an aversive outcome, the conditioning procedure also improves the strength of the association between the rewarding outcome and the correct option, as demonstrated in vertebrates (Rescorla and Wagner, 1972).
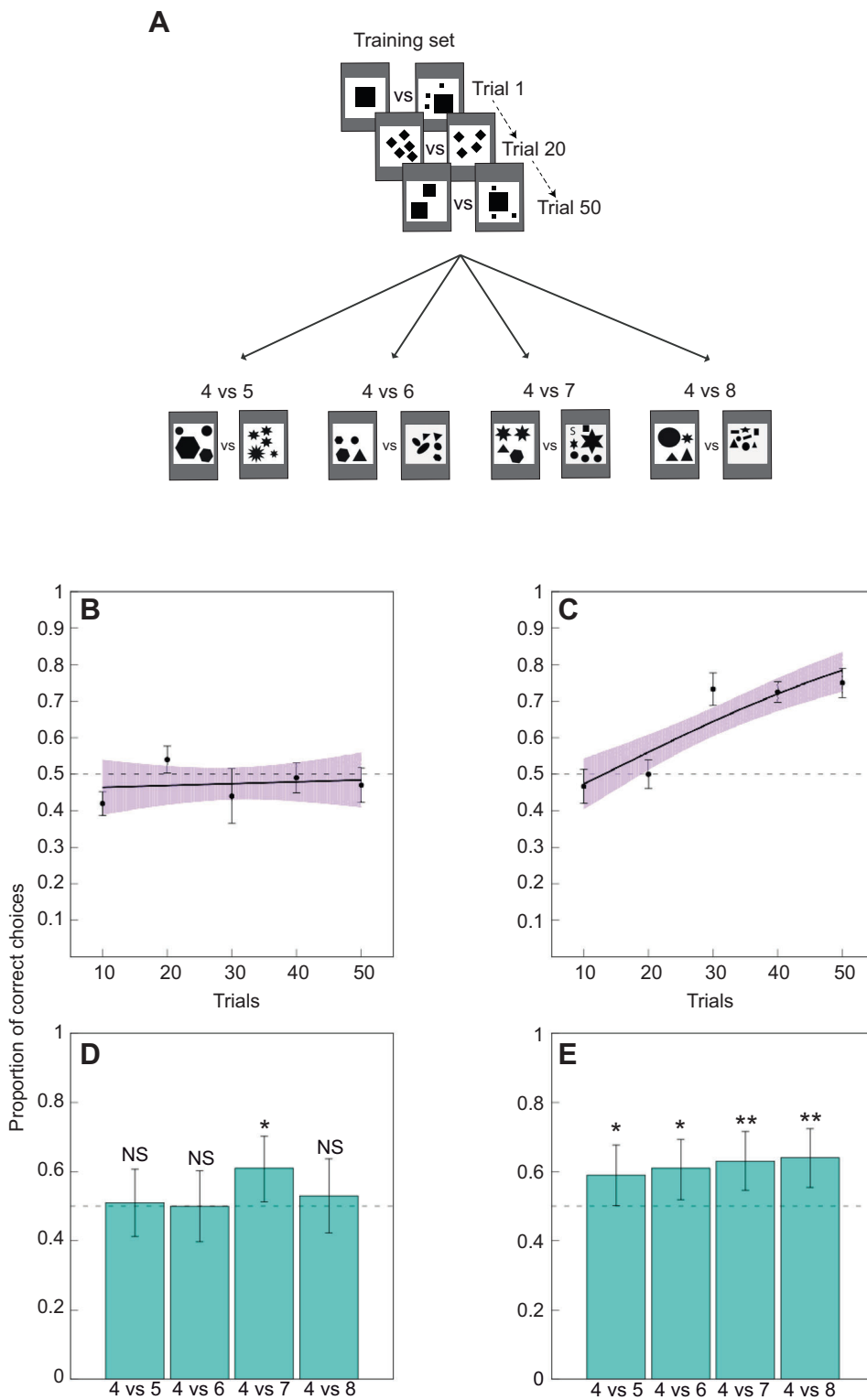
Fig. 3. Graphical representation of the method and results of group 1 and 2 training and testing. (A) Examples of possible stimulus combinations during trials and tests. (B,C) Performance during the training phase of 50 trials of either appetitive differential conditioning (B) or appetitive–aversive differential conditioning (C). (D,E) Performance during the unreinforced testing phases of group 1 (D) and group 2 (E). Data shown are means±95% confidence intervals for both groups. The dashed black line at 50% indicates chance level performance. Significance from chance level performance is indicated by asterisks (*$P \geq 0.05$, **$P \geq 0.01$). NS indicates performance that was not significantly different from chance.

Through the possible modulation of attention, we show that previous studies on animal numerical ability, specifically quantity discrimination, may have underestimated the potential numeric ability of non-human animals. In the current study, we found a difference in the results considering both the training and test phases of honeybee quantity discrimination when bees were trained with the respective procedures. Our results are supported by previous quantity discrimination studies that show varied results within the same species when different methods are employed. For example, the quantity discrimination ability of guppies has been well studied (Agrillo et al., 2012; Bisazza et al., 2014; Gatto et al., 2017; Lucon-Xiccato and Dadda, 2017; Lucon-Xiccato et al., 2017, 2015; Piffer et al., 2012, 2013), but this ability appears to be dependent on the testing procedure (Agrillo and Bisazza, 2014). Guppies tested on quantity discrimination using spontaneous choice, training, extensive training or different apparatuses (e.g. central cylindrical

restricted areas versus large square tanks) exhibit different limits of discrimination. Recording the spontaneous choices of guppies has resulted in a discrimination ability at a ratio of 0.75 for low (4 and fewer) number comparisons (Piffer et al., 2012) and 0.67 for high (4 and higher) number comparisons (Miletto Petrazzini and Agrillo, 2016); however, with extensive training, this has been extended to 0.80 with higher numbers (Bisazza et al., 2014), and even 0.83 using a recently developed method (Lucon-Xiccato et al., 2017). In the current study, we also show that different methods yield different results and quantity discrimination abilities; thus, perhaps non-human animals in other studies that are not as motivated to avoid incorrect choices may not have been pushed to their cognitive limits. Numerical processing is likely to be a valuable skill for animals operating successfully in complex environments (Gómez-Laplaza and Gerlai, 2011a,b; Lucon-Xiccato et al., 2017; Miletto Petrazzini and Agrillo, 2016; Nieder, 2017; Seguin and Gerlai, 2017), and we show that motivation, as modulated through conditioning, is critical to understanding what level of numerical ability an animal demonstrates.

Data availability
The raw choice data from tests of individual bees that support the findings of this study are available from the Dryad digital repository (Howard et al., 2019c): dryad. pm6v7nd

Supplementary information
Supplementary information available online at
http://jeb.biologists.org/lookup/doi/10.1242/jeb.205658.supplemental

References
Addessi, E., Crescimbene, L. and Visalberghi, E. (2008). Food and token quantity discrimination in capuchin monkeys (*Cebus apella*). *Anim. Cogn.* **11**, 275-282. doi:10.1007/s10071-007-0111-6

Agrillo, C. and Bisazza, A. (2014). Spontaneous versus trained numerical abilities. A comparison between the two main tools to study numerical competence in non-human animals. *J. Neurosci. Methods* **234**, 82-91. doi:10.1016/j.jneumeth.2014.04.027

Agrillo, C., Dadda, M., Serena, G. and Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Anim. Cogn.* **11**, 495-503. doi:10.1007/s10071-008-0140-9

Agrillo, C., Piffer, L., Bisazza, A. and Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PLoS ONE* **7**, e31923. doi:10.1371/journal.pone.0031923

Akre, K. L. and Johnsen, S. (2014). Psychophysics and the evolution of behavior. *Trends Ecol. Evol.* **29**, 291-300. doi:10.1016/j.tree.2014.03.007

Al Aïn, S., Giret, N., Grand, M., Kreutzer, M. and Bovet, D. (2009). The discrimination of discrete and continuous amounts in African grey parrots (*Psittacus erithacus*). *Anim. Cogn.* **12**, 145-154. doi:10.1007/s10071-008-0178-8

Avarguès-Weber, A., de Brito Sanchez, M. G., Giurfa, M. and Dyer, A. G. (2010). Aversive reinforcement improves visual discrimination learning in free-flying honeybees. *PLoS ONE* **5**, e15370. doi:10.1371/journal.pone.0015370

Avarguès-Weber, A., Dyer, A. G. and Giurfa, M. (2011). Conceptualization of above and below relationships by an insect. *Proc. R. Soc. B* **278**, 898-905. doi:10.1098/rspb.2010.1891

Avarguès-Weber, A., Dyer, A. G., Ferrah, N. and Giurfa, M. (2015). The forest or the trees: preference for global over local image processing is reversed by prior experience in honeybees. *Proc. R. Soc. B* **282**, 20142384. doi:10.1098/rspb.2014.2384

Bertamini, M., Guest, M., Vallortigara, G., Rugani, R. and Regolin, L. (2018). The effect of clustering on perceived quantity in humans (*Homo sapiens*) and in chicks (*Gallus gallus*). *J. Comp. Psychol.* **132**, 280-293. doi:10.1037/com0000114

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

Bogale, B. A., Kamata, N., Mioko, K. and Sugita, S. (2011). Quantity discrimination in jungle crows, *Corvus macrorhynchos*. *Anim. Behav.* **82**, 635-641. doi:10.1016/j.anbehav.2011.05.025

Boysen, S. T. (1988). Kanting processes in the chimpanzee: what (and who) really counts? *Behav. Brain Sci.* **11**, 580-580. doi:10.1017/S0140525X00053449

Carazo, P., Font, E., Forteza-Behrendt, E. and Desfilis, E. (2009). Quantity discrimination in Tenebrio molitor: evidence of numerosity discrimination in an invertebrate? *Anim. Cogn.* **12**, 463-470. doi:10.1007/s10071-008-0207-7

Chittka, L. and Geiger, K. (1995). Can honey bees count landmarks? *Anim. Behav.* **49**, 159-164. doi:10.1016/0003-3472(95)80163-4

Chittka, L., Dyer, A. G., Bock, F. and Dornhaus, A. (2003). Psychophysics: bees trade off foraging speed for accuracy. *Nature* **424**, 388-388. doi:10.1038/424388a

Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Curr. Direct. Psychol. Sci.* **19**, 51-57. doi:10.1177/0963721409359277

Dacke, M. and Srinivasan, M. V. (2008). Evidence for counting in insects. *Anim. Cogn.* **11**, 683-689. doi:10.1007/s10071-008-0159-y

Dehaene, S. and Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *J. Exp. Psychol. Hum. Percep. Perform.* **20**, 958-975. doi:10.1037/0096-1523.20.5.958

DeWind, N. K. and Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Front. Hum. Neurosci.* **6**, 68. doi:10.3389/fnhum.2012.00068

DeWind, N. K., Adams, G. K., Platt, M. L. and Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition* **142**, 247-265. doi:10.1016/j.cognition.2015.05.016

Dyer, A. G. (2012). The mysterious cognitive abilities of bees: why models of visual processing need to consider experience and individual differences in animal performance. *J. Exp. Biol.* **215**, 387-395. doi:10.1242/jeb.038190

Dyer, A. G. and Griffiths, D. W. (2012). Seeing near and seeing far; behavioural evidence for dual mechanisms of pattern vision in the honeybee (*Apis mellifera*). *J. Exp. Biol.* **215**, 397-404. doi:10.1242/jeb.060954

Dyer, A. G., Neumeyer, C. and Chittka, L. (2005). Honeybee (*Apis mellifera*) vision can discriminate between and recognise images of human faces. *J. Exp. Biol.* **208**, 4709-4714. doi:10.1242/jeb.01929

Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. New York: Chapman and Hall/CRC.

Fechner, G. (1965). *Elements of Psychophysics*. New York: Rinehart & Winston.

Feigenson, L., Dehaene, S. and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* **8**, 307-314. doi:10.1016/j.tics.2004.05.002

Gabor, V. and Gerken, M. (2018). Study into long-term memory of a complex learning task in Shetland ponies (*Equus caballus*). *Appl. Anim. Behav. Sci.* **198**, 60-66. doi:10.1016/j.applanim.2017.09.010

Garland, A., Low, J. and Burns, K. C. (2012). Large quantity discrimination by North Island robins (*Petroica longipes*). *Anim. Cogn.* **15**, 1129-1140. doi:10.1007/s10071-012-0537-3

Gatto, E., Lucon-Xiccato, T., Savaşçı, B. B., Dadda, M. and 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

Gazes, R. P., Billas, A. R. and Schmitt, V. (2018). Impact of stimulus format and reward value on quantity discrimination in capuchin and squirrel monkeys. *Learn. Behav.* **46**, 89-100. doi:10.3758/s13420-017-0295-9

Giurfa, M. (2004). Conditioning procedure and color discrimination in the honeybee *Apis mellifera*. *Naturwissenschaften* **91**, 228-231. doi:10.1007/s00114-004-0530-z

Giurfa, M., Hammer, M., Stach, S., Stollhoff, N., Müller-Deisig, N. and Mizyrycki, C. (1999). Pattern learning by honeybees: conditioning procedure and recognition strategy. *Anim. Behav.* **57**, 315-324. doi:10.1006/anbe.1998.0957

Gómez-Laplaza, L. M. and Gerlai, R. (2011a). Can angelfish (*Pterophyllum scalare*) count? Discrimination between different shoal sizes follows Weber's law. *Anim. Cogn.* **14**, 1-9. doi:10.1007/s10071-010-0337-6

Gómez-Laplaza, L. M. and Gerlai, R. (2011b). Spontaneous discrimination of small quantities: shoaling preferences in angelfish (*Pterophyllum scalare*). *Anim. Cogn.* **14**, 565-574. doi:10.1007/s10071-011-0392-7

Gross, H. J., Pahl, M., Si, A., Zhu, H., Tautz, J. and Zhang, S. (2009). Number-based visual generalisation in the honeybee. *PLoS ONE* **4**, e4263. doi:10.1371/journal.pone.0004263

Hanus, D. and Call, J. (2007). Discrete quantity judgments in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*): the effect of

presenting whole sets versus item-by-item. *J. Comp. Psychol.* **121**, 241-249. doi:10.1037/0735-7036.121.3.241

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Stuart-Fox, D. and Dyer, A. G. (2017). Perception of contextual size illusions by honeybees in restricted and unrestricted viewing conditions. *Proc. R. Soc. B* **284**, 20172278. doi:10.1098/rspb.2017.2278

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D. and Dyer, A. G. (2018). Numerical ordering of zero in honey bees. *Science* **360**, 1124-1126. doi:10.1126/science.aar4975

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D. and Dyer, A. G. (2019a). Numerical cognition in honeybees enables addition and subtraction. *Sci. Adv.* **5**, eaav0961. doi:10.1126/sciadv.aav0961

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D. and Dyer, A. G. (2019b). Symbolic representation of numerosity by honeybees (*Apis mellifera*): matching characters to small quantities. *Proc. R. Soc. B* **286**, 20190238. doi:10.1098/rspb.2019.0238

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D. and Dyer, A. G. (2019c). Data from: Surpassing the subitizing threshold: appetitive-aversive conditioning improves discrimination of numerosities in honeybees. *Dryad Digital Repository.* https://doi.org/10.5061/dryad.pm6v7nd

Irie-Sugimoto, N., Kobayashi, T., Sato, T. and Hasegawa, T. (2009). Relative quantity judgment by Asian elephants (*Elephas maximus*). *Anim. Cogn.* **12**, 193-199. doi:10.1007/s10071-008-0185-9

Jaakkola, K., Fellner, W., Erb, L., Rodriguez, M. and Guarino, E. (2005). Understanding of the concept of numerically "less" by bottlenose dolphins (*Tursiops truncatus*). *J. Comp. Psychol.* **119**, 296-303. doi:10.1037/0735-7036.119.3.296

Jevons, W. S. (1871). The power of numerical discrimination. *Nature* **3**, 281-282. doi:10.1038/003281a0

Kaufman, E. L., Lord, M. W., Reese, T. W. and Volkmann, J. (1949). The discrimination of visual number. *Am. J. Psychol.* **62**, 498-525. doi:10.2307/1418556

Kelly, E. M. K. (2016). Counting on your friends: The role of social environment on quantity discrimination. *Behav. Process.* **128**, 9-16. doi:10.1016/j.beproc.2016.03.019

Lucon-Xiccato, T. and Dadda, M. (2017). Individual guppies differ in quantity discrimination performance across antipredator and foraging contexts. *Behav. Ecol. Sociobiol.* **71**, 13. doi:10.1007/s00265-016-2231-y

Lucon-Xiccato, T., Miletto Petrazzini, M. E., Agrillo, C. and 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

Lucon-Xiccato, T., Dadda, M., Gatto, E. and Bisazza, A. (2017). Development and testing of a rapid method for measuring shoal size discrimination. *Anim. Cogn.* **20**, 149-157. doi:10.1007/s10071-016-1050-x

Miletto Petrazzini, M. E. and Agrillo, C. (2016). Turning to the larger shoal: are there individual differences in small- and large-quantity discrimination of guppies? *Ethol. Ecol. Evol.* **28**, 211-220. doi:10.1080/03949370.2015.1029011

Miletto Petrazzini, M. E., Bertolucci, C. and Foà, A. (2018). Quantity discrimination in trained lizards (*Podarcis sicula*). *Front. Psychol.* **9**, 274. doi:10.3389/fpsyg.2018.00274

Nieder, A. (2017). Number faculty is rooted in our biological heritage. *Trends Cogn. Sci.* **21**, 403-404. doi:10.1016/j.tics.2017.03.014

Piazza, M., Mechelli, A., Butterworth, B. and Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage* **15**, 435-446. doi:10.1006/nimg.2001.0980

Piffer, L., Agrillo, C. and Hyde, D. C. (2012). Small and large number discrimination in guppies. *Anim. Cogn.* **15**, 215-221. doi:10.1007/s10071-011-0447-9

Piffer, L., Miletto Petrazzini, M. E. and Agrillo, C. (2013). Large number discrimination in newborn fish. *PLoS ONE* **8**, e62466. doi:10.1371/journal.pone.0062466

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, Vol. 2 (ed. A. H. Black and W. T. Prokasy), pp. 64-99. Appleton Century Crofts.

Riggs, K. J., Ferrand, L., Lancelin, D., Fryziel, L., Dumur, G. and Simpson, A. (2006). Subitizing in tactile perception. *Psychol. Sci.* **17**, 271-272. doi:10.1111/j.1467-9280.2006.01696.x

Seguin, D. and Gerlai, R. (2017). Zebrafish prefer larger to smaller shoals: analysis of quantity estimation in a genetically tractable model organism. *Anim. Cogn.* **20**, 813-821. doi:10.1007/s10071-017-1102-x

Simons, D. and Langheinrich, D. (1982). What is magic about the magical number four? *Psychol. Res.* **44**, 283-294. doi:10.1007/BF00308426

Srinivasan, M. V. and Lehrer, M. (1988). Spatial acuity of honeybee vision and its spectral properties. *J. Comp. Physiol. A* **162**, 159-172. doi:10.1007/BF00606081

Stach, S. and Giurfa, M. (2005). The influence of training length on generalization of visual feature assemblies in honeybees. *Behav. Brain Res.* **161**, 8-17. doi:10.1016/j.bbr.2005.02.008

Stancher, G., Rugani, R., Regolin, L. and Vallortigara, G. (2015). Numerical discrimination by frogs (*Bombina orientalis*). *Anim. Cogn.* **18**, 219-229. doi:10.1007/s10071-014-0791-7

Starkey, P. and Cooper, R. G., Jr. (1995). The development of subitizing in young children. *Br. J. Dev. Psychol.* **13**, 399-420. doi:10.1111/j.2044-835X.1995.tb00688.x

Tomonaga, M. and Matsuzawa, T. (2002). Enumeration of briefly presented items by the chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*). *Anim. Learn. Behav.* **30**, 143-157. doi:10.3758/BF03192916

Tornick, J. K., Callahan, E. S. and Gibson, B. M. (2015). An investigation of quantity discrimination in Clark's nutcrackers (*Nucifraga columbiana*). *J. Comp. Psychol.* **129**, 17-25. doi:10.1037/a0037863

Trick, L. M. and Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol. Rev.* **101**, 80-102. doi:10.1037/0033-295X.101.1.80

Vallortigara, G. (2017). 2.2 An animal's sense of number. In *The Nature and Development of Mathematics: Cross Disciplinary Perspectives on Cognition, Learning and Culture* (ed. J. Adams, P. Barmby and A. Mesoudi), pp. 43-65. New York: Routledge.

Weber, E. H. (1978). *EH Weber: The Sense of Touch*. Academic Press.
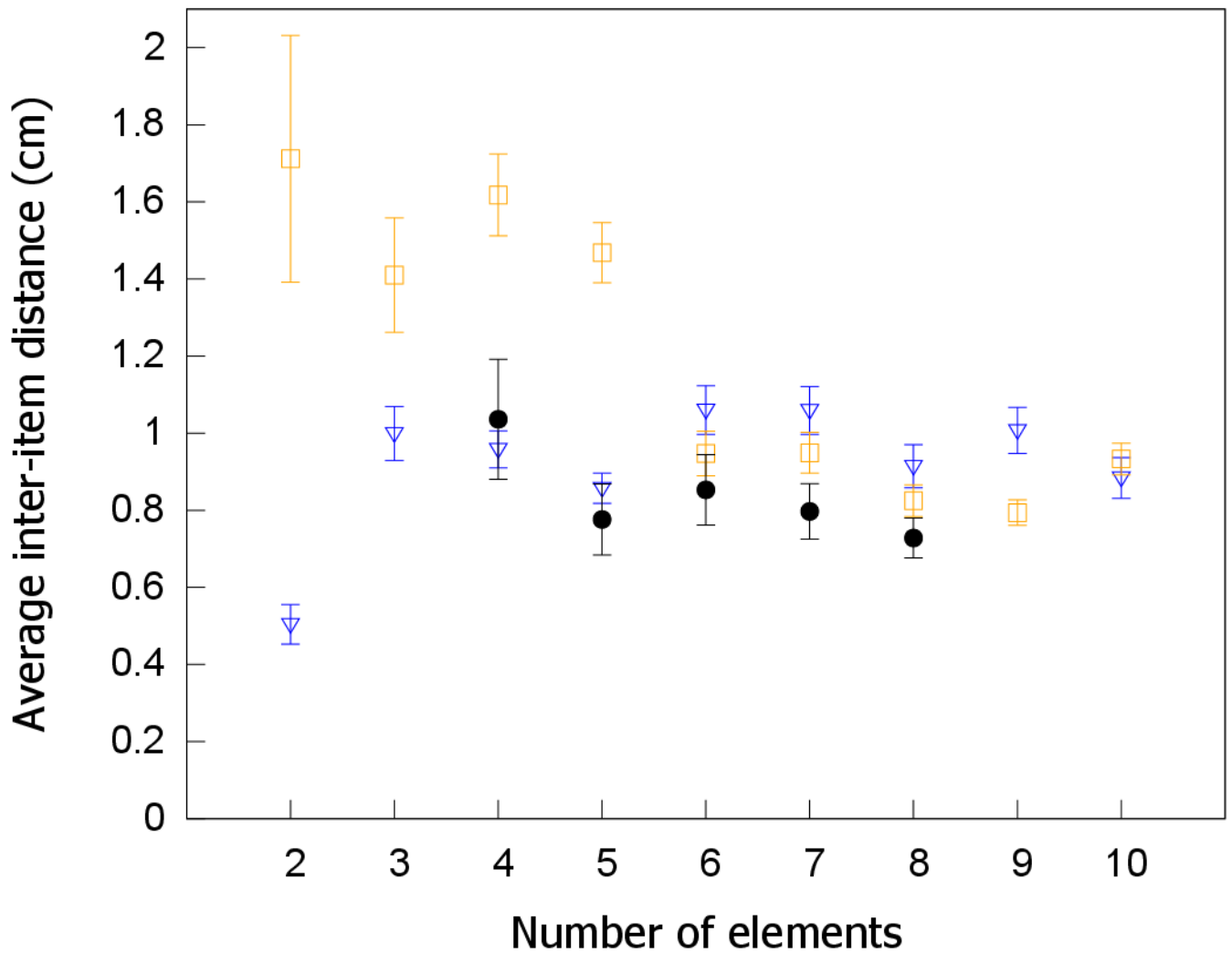
Figure S1: A plot showing the average inter-item distance of elements (cm) in set 1 (equal overall surface area of elements) shown in blue (triangles), set 2 (equal element area) shown in orange (squares), and set 3 (test set) shown in black (circles). The data shown here is the average inter-item distance of neighbouring elements for each set ± the standard error of the mean. Data was collected by measuring the shortest distance between all neighbouring elements in each stimulus using the program ImageJ. Two elements were considered neighbours if less than 50 % of the element was occluded by another element.
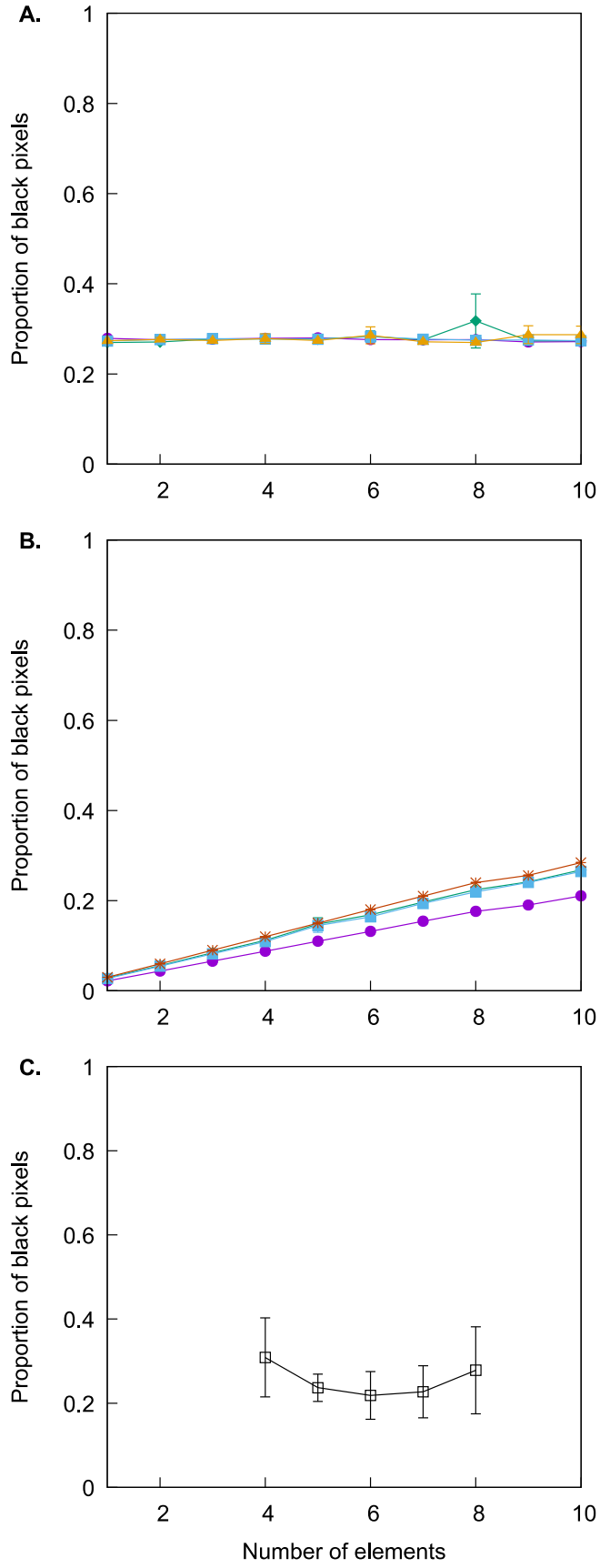
Figure S2: Mean proportion of black pixels of each element shape in stimuli sets with different number of elements with same surface area (panel A, top), same element size (panel B, middle), and the test stimuli (panel C, bottom). For each stimulus, the proportion of black pixels was calculated as the number of black pixels (points) making up the various elements present on each stimulus divided over the total number of pixels making each image (29,241 pixels). In panel A and panel B, purple circles represent the stimuli containing circle-shaped elements, green diamonds represent stimuli containing diamond-shaped elements, and the blue squares represent stimuli containing square-shaped elements. Yellow triangles in panel A represent stimuli containing triangle-shaped elements, and red asterisks in panel B represent stimuli containing star-shaped elements. Panel C shows the mean proportion of black pixels for the test stimuli (black line with open square markers). Error bars represents one standard deviation from the mean in all cases.