iScience, Volume *23*

Supplemental Information

Transfer from Number to Size Reveals Abstract Coding of Magnitude in Honeybees

Maria Bortot, Gionata Stancher, and Giorgio Vallortigara

SUPPLEMENTAL ITEMS

FIGURES

Figure S1. Numerical contrast 2 *vs.* 3, Related to Figure 1a.

The 90 stimuli pairs displaying the 2 *vs.* 3 numerical contrast. Stimuli are divided by shape categories: circles, diamond and squares. Within each shape, in one quarter of the pairs the cumulative surface area was matched to 100% whereas in the second quarter it was not controlled. In the third and the fourth quarter the contour length was matched to 100% and not controlled, respectively. Moreover, within each shape, half of the stimuli was controlled for the convex hull and the other half for the density of the elements.

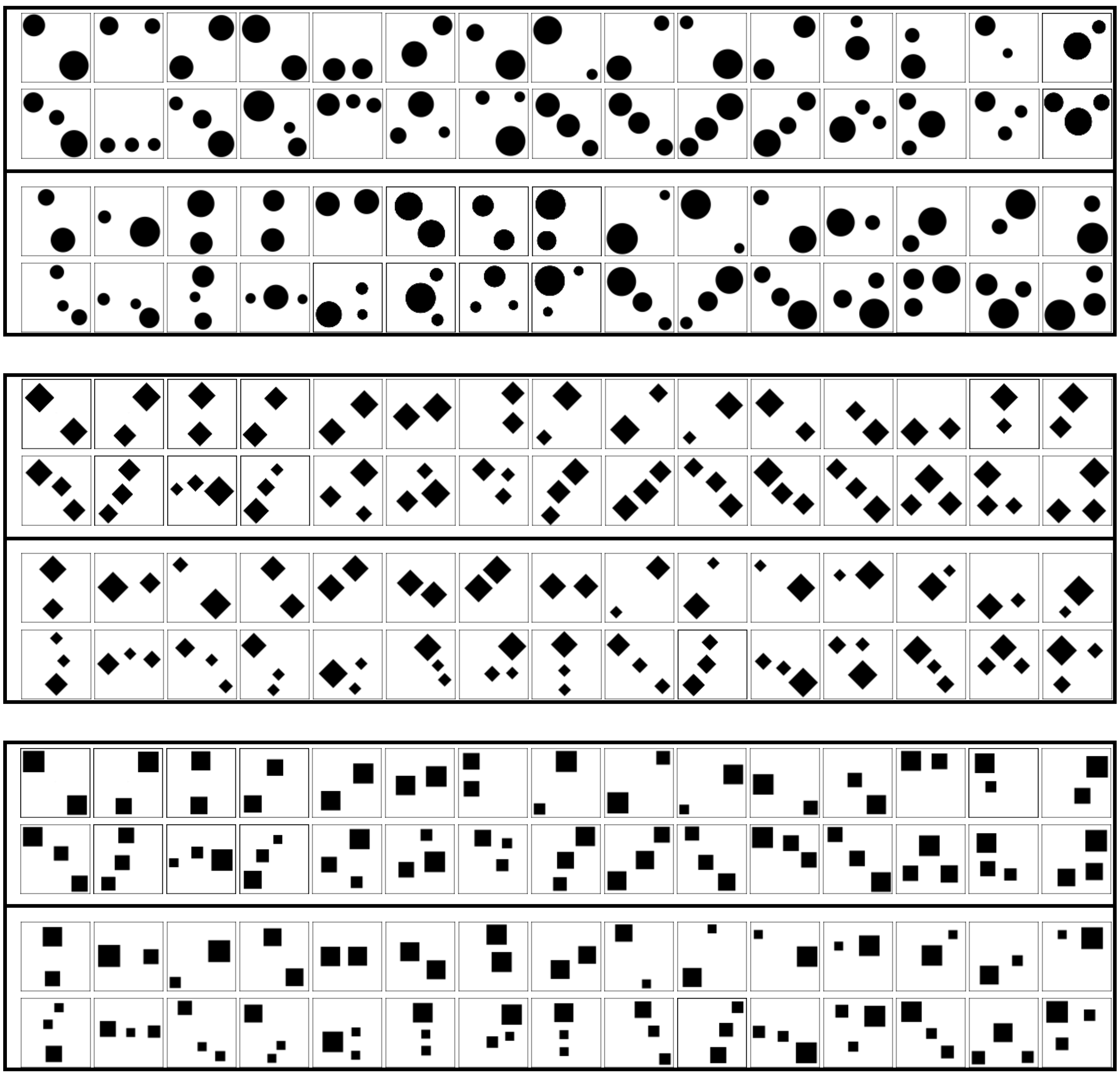


Figure S2. Numerical contrast 2 *vs.* 4, Related to Figure 1a.

The 90 stimuli pairs displaying the 2 *vs.* 4 numerical contrast. Stimuli are divided by shape categories: circles, diamond and squares. Within each shape, in one quarter of the pairs the cumulative surface area was matched to 100% whereas in the second quarter it was not controlled. In the third and the fourth quarter the contour length was matched to 100% and not controlled, respectively. Moreover, within each shape, half of the stimuli was controlled for the convex hull and the other half for the density of the elements.

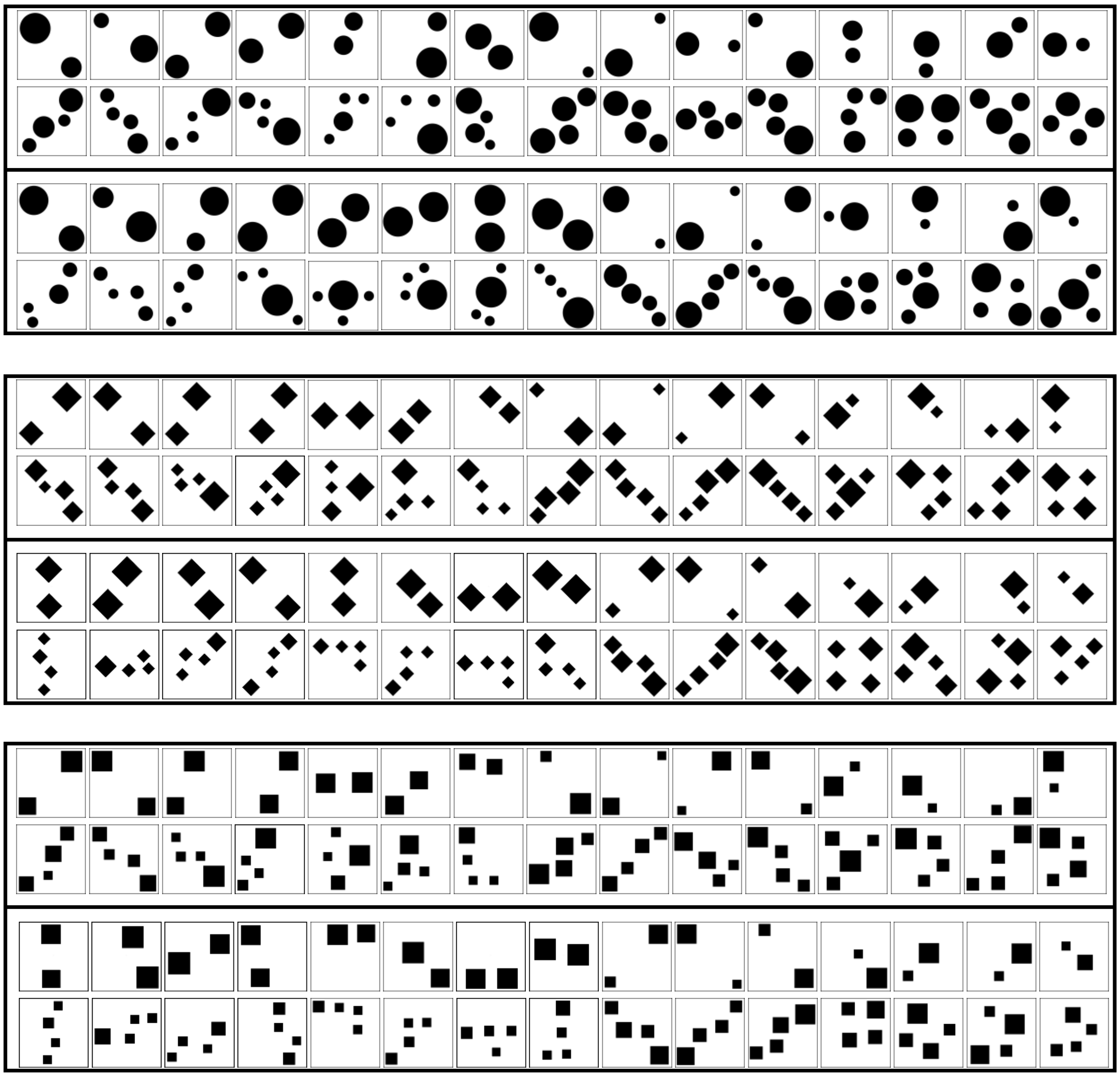


Figure S3. Numerical contrast 4 *vs.* 6, Related to Figure 1a.

The 90 stimuli pairs displaying the 4 *vs.* 6 numerical contrast. Stimuli are divided by shape categories: circles, diamond and squares. Within each shape, in one quarter of the stimuli the cumulative surface area was matched to 100% whereas in the second quarter it was not controlled. In the third and the fourth quarter the contour length was matched to 100% and not controlled, respectively. Moreover, within each shape, half of the stimuli was controlled for the convex hull and the other half for the density of the elements.

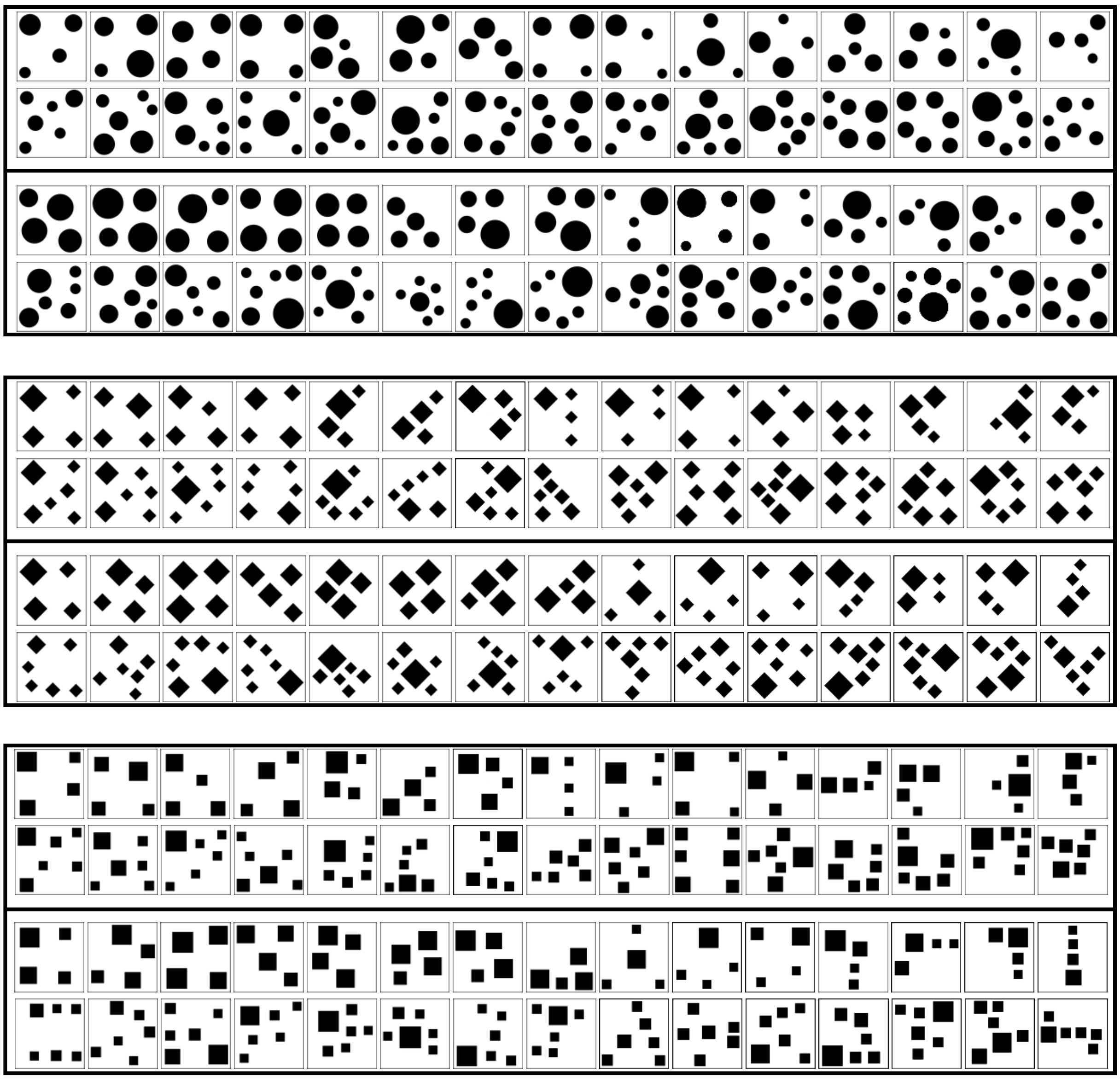


Figure S4. Numerical contrast 4 *vs.* 8, Related to Figure 1a.

The 90 stimuli pairs displaying the 4 *vs.* 8 numerical contrast. Stimuli are divided by shape categories: circles, diamond and squares. Within each shape, in one quarter of the pairs the cumulative surface area was matched to 100% whereas in the second quarter it was not controlled. In the third and the fourth quarter the contour length was matched to 100% and not controlled, respectively. Moreover, within each shape, half of the stimuli was controlled for the convex hull and the other half for the density of the elements.

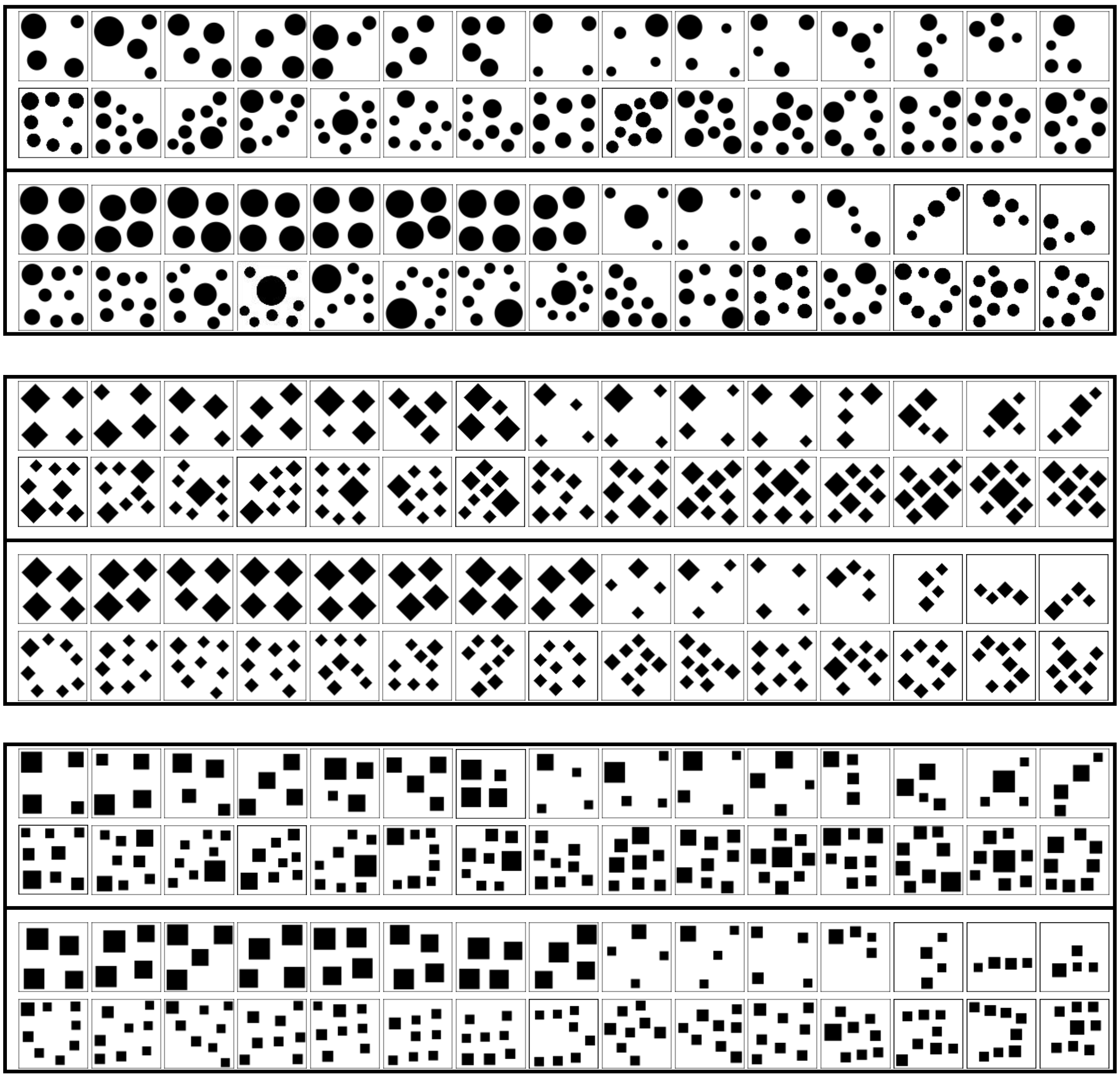


Figure S5. Size contrast, Related to Figure 1b.

The stimuli pairs used during the *size transfer* test. The stimuli consisted of two pairs of novel shapes (i.e., the shape that was not presented in the training phase) with sizes that differed by a ratio of either 0.5 or 0.67, depending on the numerical training previously completed by each subject. Within each pair, the two arrays had the same number and disposition of elements.

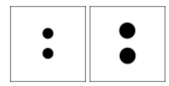
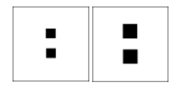
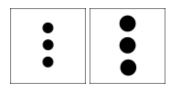
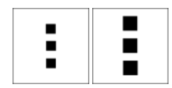
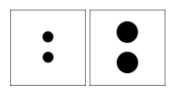
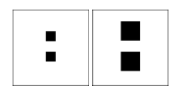
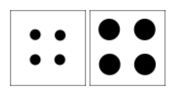
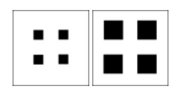
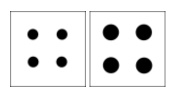
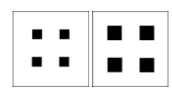
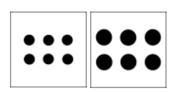
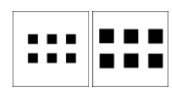
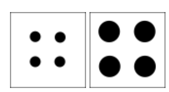
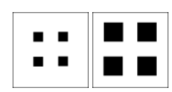
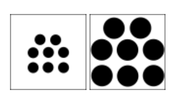
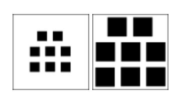
Numerical contrast	Correct numerosity during training	Stimuli *size transfer* test	
		Circles	Squares
2 *vs.* 3	2 elements		
2 *vs.* 3	3 elements		
2 *vs.* 4	2 elements		
2 *vs.* 4	4 elements		
4 *vs.* 6	4 elements		
4 *vs.* 6	6 elements		
4 *vs.* 8	4 elements		
4 *vs.* 8	8 elements		

Figure S6. Original scoring, Related to Figure 1a,b.

The original scoring of the test phase of 30 subjects (*** $P < 0.001$, Analysis of variance (ANOVA)).

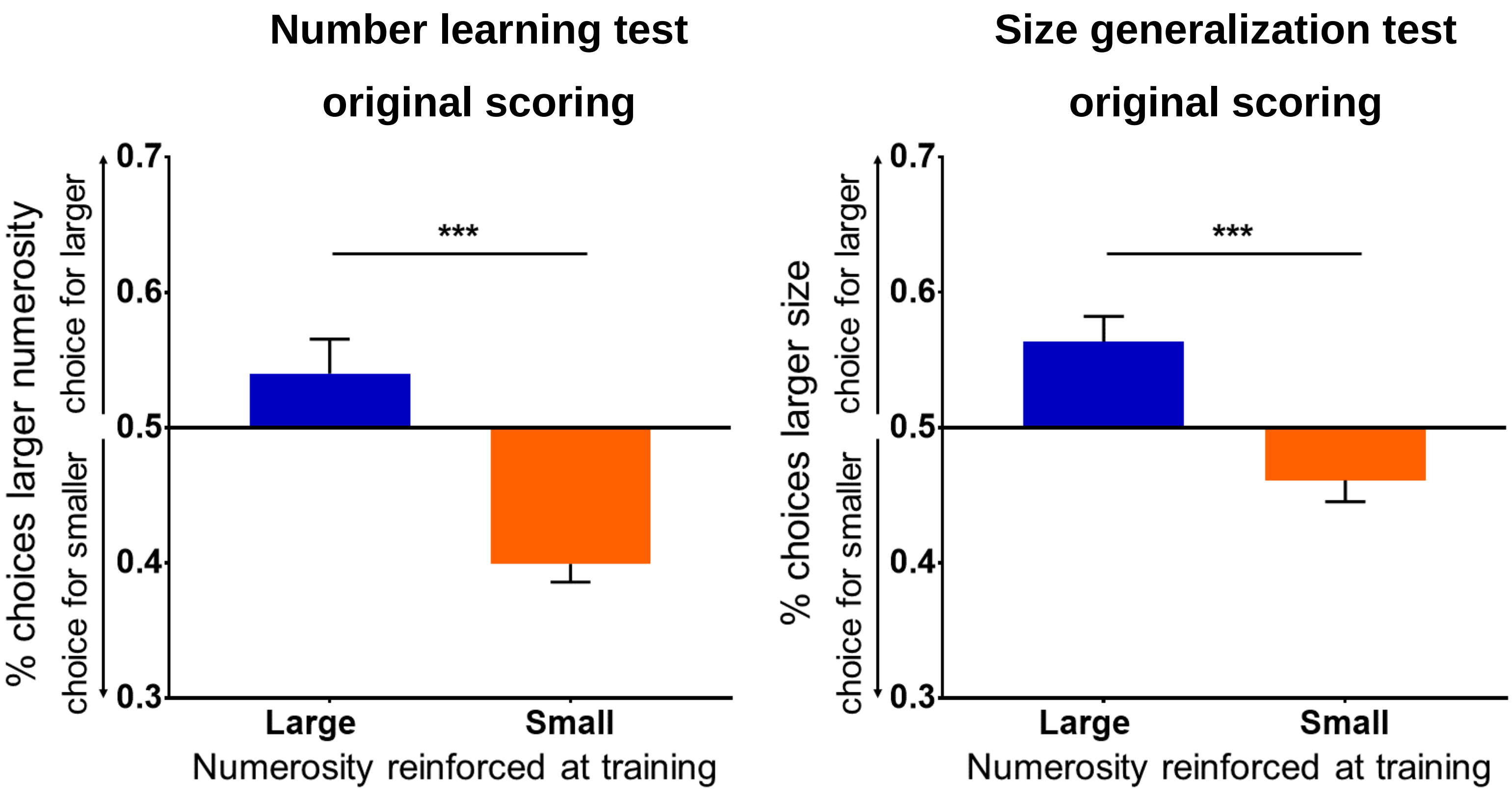
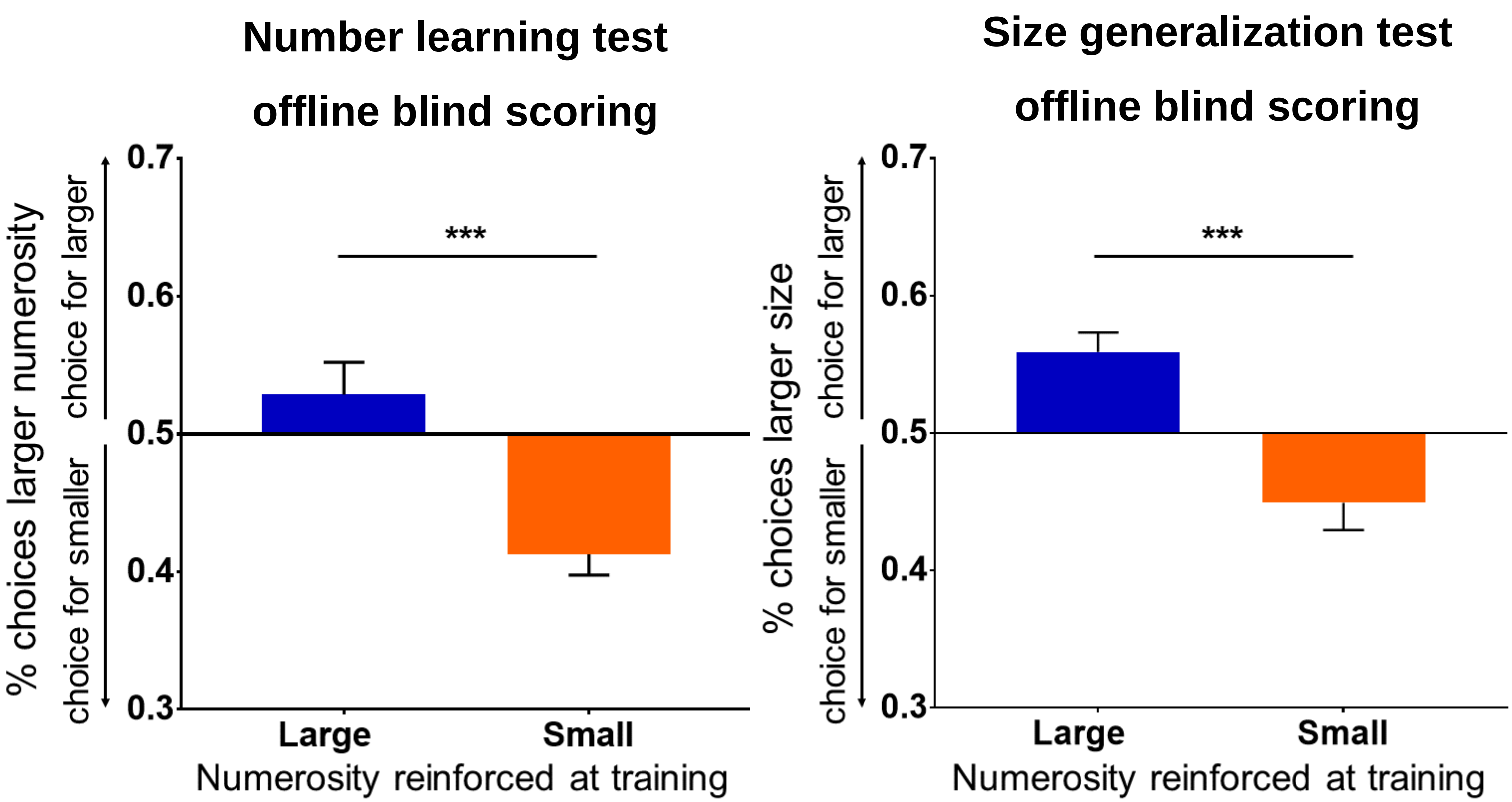


Figure S7. Offline blind scoring, Related to Figure 1a,b.

The offline blind scoring of the test phase of 30 subjects (*** *P* < 0.001, Analysis of variance (ANOVA))



TRANSPARENT METHODS

Experiments were performed during the Summer 2019 at SperimentArea, a field station run by the local Natural History Museum, in Rovereto (North of Italy). Thirty-two free-flying honeybees (*Apis mellifera*) were trained singly to fly into a wooden Y-maze (Fig. 1c). This sample size is common in experiments on free-flying honeybees' visual learning abilities (Bortot et al., 2019; Howard et al., 2019; Howard et al., 2018), due to the nature of the free-flight condition that do not allow a control by the experimenter on the decision of the single bee to come back freely to the apparatus. One half of the bees were trained with a 0.5 ratio (N=16) and the other half with a 0.67 ratio (N=16). In the 0.5 ratio one half of the subjects was tested with a 2 *vs.* 4 comparison and the other half with a 4 *vs.* 8 comparison; in the 0.67 ratio one half of the subjects was tested with a 2 *vs.* 3 comparison and the other half with a 4 *vs.* 6 comparison.

The stimuli consisted of black elements, either squares, diamonds or dots on a white squared-shape background (8 cm x 8 cm) located at 15 cm distance from the decision chamber (Fig. 1c). The stimuli size ranged from 1.12 cm to 3.56 cm (diameter of dots) and from 1 cm to 2.5 cm (side of squares and diamonds). A total of 30 couples of stimuli were used for each shape (i.e., squares, diamonds and dots). The spatial disposition and the size of the elements were varied among trials to prevent the use of non-numerical cues. In order to control for the continuous variables that covary with numerosity (e.g., area, contour length, density), we adopted a procedure previously used in other studies on numerical abilities of bees (Bortot et al., 2019). Within each shape, in one quarter of the stimuli (N=7) the cumulative surface area was matched to 100%, whereas in the second quarter (N=8) was not controlled (i.e., the ratio between the cumulative surface area within each pair was congruent with the numerical ratio: 0.5 in 2 *vs.* 4 and 4 *vs.* 8; 0.67 in 2 *vs.* 3 and 4 *vs.* 6). In addition, half of the stimuli was controlled for the convex hull and the other half for the density of the elements. Furthermore, in the third and fourth quarter of the stimuli, the cumulative contour length was matched to 100% (N=8) and not controlled (N=7), respectively, following the same logic. Again, within each shape, half of the stimuli were controlled for the convex hull and the other half for the density of the elements. The control of these variables was performed for each shape (Fig. S1, S2, S3, S4).

During the training phase, half of the bees (N=16) were presented with squares and diamonds, whereas the other half (N=16) was presented with diamonds and dots. Thus, in the training phase bees were presented with 60 couple of stimuli in random order differing in shape, spatial disposition, size of the element and combinations of controlled continuous variable, and only the numerical information was kept constant. The stimuli used in the *number learning* test, were taken from the training sample of stimuli with the area matched to 100%. In the *size transfer* test, stimuli consisted of two pairs of novel shapes (i.e., the shape that was not presented in the training phase) having sizes that differed by a ratio of either 0.5 or 0.67, depending on the numerical training previously completed by each subject. Within each pair, the two arrays had the same number and disposition of elements. In particular, the number of elements presented was equal to the numerosity reinforced during the training phase (e.g., bees trained to select 2 elements over 4 elements during the training phase, were then presented with a 2 *vs*. 2 comparison where one group of 2 elements had the double size of the other group of 2 elements) (Fig. S5).

The experimental procedure comprised a pre-training phase followed by a training and tests phase. All the phases were completed by all subjects in 1 or 2 consecutive days. During the pre-training phase, each bee was individually habituated to fly inside the apparatus and to collect food by landing on two grey poles placed in both arms, in the absence of visual stimuli. In the training phase, four different numerical comparisons (ratio 0.5: 2 *vs.* 4, 4 *vs.* 8; ratio 0.67: 2 *vs.* 3, 4 *vs.* 6) were presented to each independent group, separately. Within each group, half of the subjects was trained to select the smaller numerosity in the comparisons (either 2 or 4), whereas the other half was trained to choose the larger numerosity in the comparison (either 3, 4, 6 or 8) in order to get the food reward. During this phase, an appetitive-aversive conditioning paradigm was used: the correct numerosity was always associated with the food (0.88 M of sucrose solution) whereas the incorrect numerosity was always associated with a bitter 60 mM quinine solution, used as punishment (Avarguès-Weber et al., 2018). The use of this appetitive-aversive conditioning has been shown to improve the ability of bees to discriminate between numerosities (Howard et al., 2019). Each subject had to complete 60 consecutive trials of training. The stimuli were presented in a pseudo-random sequence (i.e., the correct/incorrect stimulus was never presented for more than two consecutive times on the same side).

Once completed the training phase, honeybees started the test phase. During this phase, two non-reinforced tests were presented: a *number learning* test and a *size transfer* test. Each test was presented twice to counterbalance the position of the correct array and avoid side preferences. The tests lasted 1 minute during which the number of choices (i.e., direct contact made with a body part, either the antennae or legs, on one of the two grey poles placed in front of each stimulus) made by the subjects were counted online. In the *number learning* test, bees were presented with the same numerical comparisons and shapes used during the training but in the absence of any reward. In the *size transfer* test bees were exposed to the novel stimuli displaying only the size information (even in this case without any reward).

In the test phase, the percentage of choices for the larger numerosity and larger size was calculated for each subject and analyzed, giving rise one single value per bee to exclude pseudo-replication. The data were checked for normality (Shapiro-Wilk normality test: $W = 0.98$, $P > 0.05$) and homoscedasticity (Levene's test: $P > 0.05$) and then analyzed with parametric statistical tests. An analysis of variance was performed with ratio (0.5 and 0.67), type of training (smaller *vs.* larger as positive) and type of test (*number learning* test *vs. size generalization* test) as factors. The effect of the numerical comparisons, as factor nested in ratio, was analyzed with a nested factorial Anova. The omega-squared effect size of any significant results in the Anova analysis was reported. The proportion of choices for the correct numerosity during the *number learning* test and the proportion of choices for the congruent size during the *size generalization* test were calculated for each subject and analyzed with a two-tailed one-sample *t*-test. The Cohen's *d* effect size of any significant result was also reported.

Analyses Blind Video Coding

We performed an offline scoring in blind condition for the videos of the test phase and compared them with the original analysis (due to corruption of two videos, 30 subjects were used in both conditions). The results are shown in the Figure S6 (original scorings) and Figure S7 (offline blind scoring) below together with statistical analyses and confirmed the original analyses.

For each group – *offline blind scoring* and *original* scoring - we calculate the percentage of choices for the larger numerosity during the *number learning* test and for the larger size during the *size generalization* test for each subject and analyzed. The data were checked for normality (*offline blind scoring*: Shapiro-Wilk normality test: W: 0.99, $P > 0.05$; *original scoring*: Shapiro-Wilk normality test: W: 0.98, $P > 0.05$) and homoscedasticity (*offline blind scoring*: Levene's test: $P >$

0.05; *original scoring*: Levene's test: P > 0.05) and then analyzed with parametric statistical tests. An analysis of variance was performed with ratio (0.5 and 0.67), type of training (smaller *vs.* larger as positive) and type of test (*number learning* test *vs. size generalization* test) as factors. The effect of the numerical comparisons, as factors nested in ratio, was analyzed with a nested factorial Anova. The omega-squared effect size of any significant results was reported.

Results

Offline blind scoring

An analysis of variance revealed a significant main effect of the type of training (smaller *vs.* larger numerosity as positive; $F_{(1, 44)} = 35.9$, $P < 0.001$, $\omega^2 = 0.371$) but not of the type of test (*number learning* test *vs. size generalization* test; $F_{(1, 44)} = 3.5$, $P = 0.069$) and of the ratio (0.5 *vs.* 0.67; $F_{(1, 44)} = 0.19$, $P > 0.05$).

No significant interactions were observed (ratio x numerical comparisons: $F_{(2, 44)} = 1.7$, $P > 0.05$; ratio x type of training: $F_{(1, 44)} = 2.2$, $P > 0.05$; ratio x type of test: $F_{(1, 44)} = 1.06$, $P > 0.05$; type of training x type of test: $F_{(1, 44)} = 0.09$, $P > 0.05$; ratio x type of training x type of test: $F_{(1, 44)} = 1.2$, $P > 0.05$; ratio x type of training x numerical comparisons: $F_{(2, 44)} = 0.04$, $P > 0.05$; ratio x type of test x numerical comparisons: $F_{(2, 44)} = 0.1$, $P > 0.05$; ratio x type of training x type of test x numerical comparisons: $F_{(2, 44)} = 0.5$, $P > 0.05$).

Original scoring

An analysis of variance revealed a significant main effect of the type of training (smaller *vs.* larger numerosity as positive; $F_{(1, 44)} = 41.6$, $P < 0.001$, $\omega^2 = 0.389$) and of the type of test (*number* *learning* test *vs. size generalization* test; $F_{(1, 44)} = 5.1$, $P = 0.029$, $\omega^2 = 0.039$) but not of the ratio (0.5 *vs.* 0.67; $F_{(1, 44)} < 0.0005$, $P > 0.05$).

No significant interactions were observed (ratio x numerical comparisons: $F_{(2, 44)} = 2.2$, $P > 0.05$; ratio x type of training: $F_{(1, 44)} = 2.3$, $P > 0.05$; ratio x type of test: $F_{(1, 44)} = 0.4$, $P > 0.05$; type of training x type of test: $F_{(1, 44)} = 1.1$, $P > 0.05$; ratio x type of training x type of test: $F_{(1, 44)} = 0.6$, $P > 0.05$; ratio x type of training x numerical comparisons: $F_{(2, 44)} = 0.8$, $P > 0.05$; ratio x type of test x numerical comparisons: $F_{(2, 44)} = 1.09$, $P > 0.05$; ratio x type of training x type of test x numerical comparisons: $F_{(2, 44)} = 0.06$, $P > 0.05$).

References

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*(10).

Bortot, M., Agrillo, C., Avarguès-Weber, A., Bisazza, A., Miletto Petrazzini, M.E. and Giurfa, M., 2019. Honeybees use absolute rather than relative numerosity in number discrimination. *Biology Letters*, *15*(6), p.20190138.

Howard, S.R., Avarguès-Weber, A., Garcia, J.E., Greentree, A.D. and Dyer, A.G., 2019. Numerical cognition in honeybees enables addition and subtraction. *Science advances*, *5*(2), p.eaav0961.

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*(6393), pp.1124-1126.

Howard, S.R., Avarguès-Weber, A., Garcia, J.E., Greentree, A.D. and Dyer, A.G., 2019. Surpassing the subitizing threshold: appetitive–aversive conditioning improves discrimination of numerosities in honeybees. *Journal of Experimental Biology*, *222*(19), p.jeb205658.