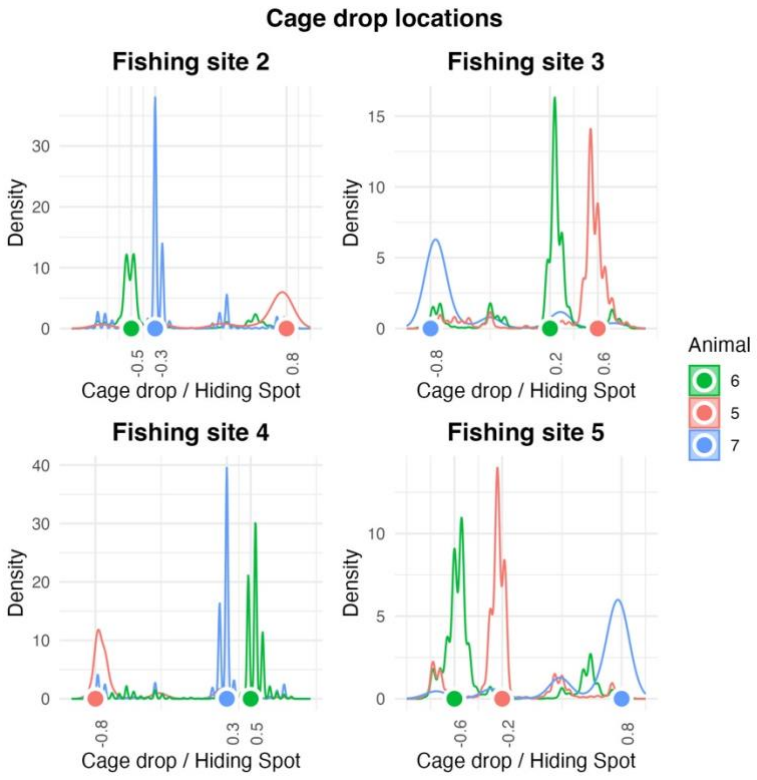
Supplementary Materials

Cage drop Locations

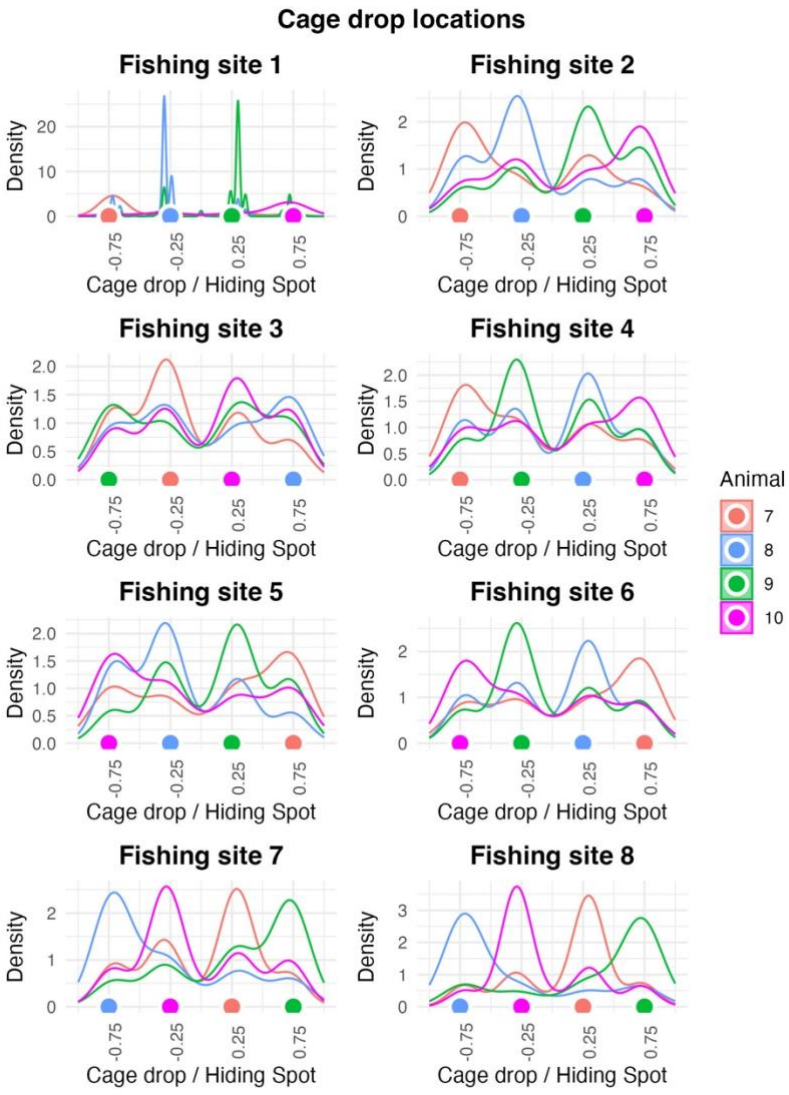
Below are three figures, each corresponding to an experiment described in the main text, showing histograms of cage drop locations for each animal in each context. Notably, these figures represent data from the generalization test rounds, where participants received no feedback and relied solely on prior learning for generalization.

Fig. S1. Cage drop locations of Experiment 1



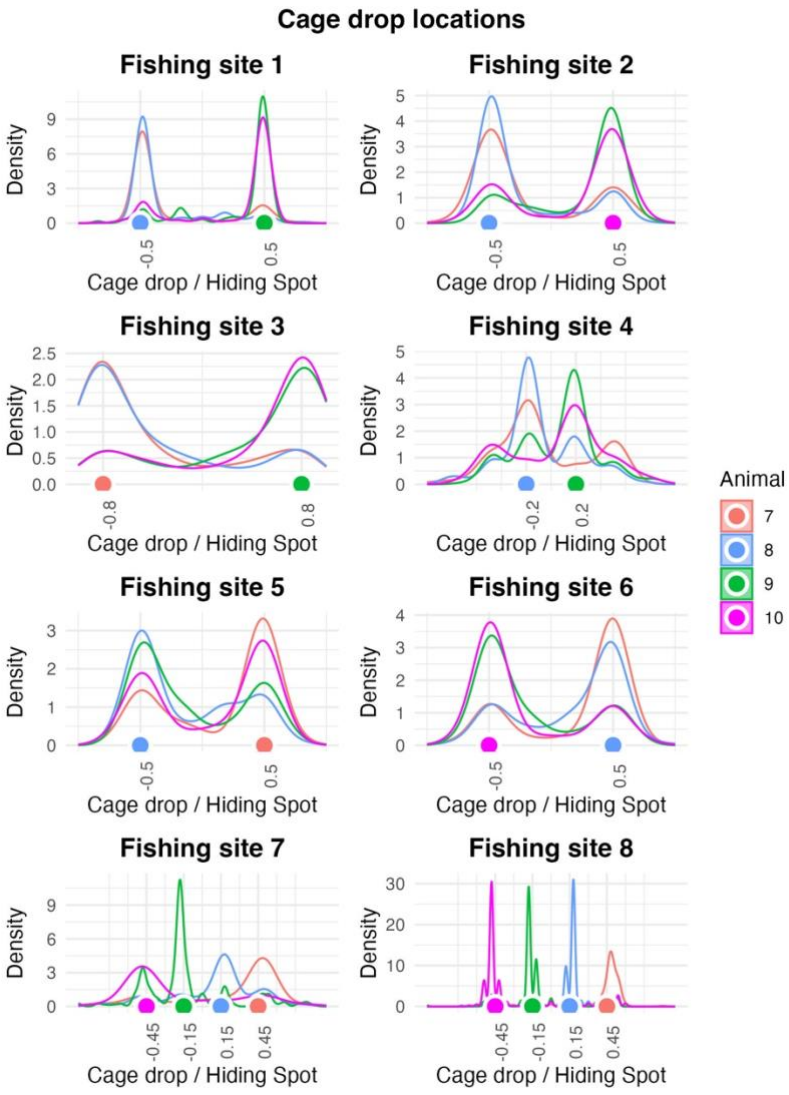
Histograms are shown of all individual cage drops in the generalization test rounds across participants. Different colors are used for the different animals. The dots represent the actual hiding spots of the animal. Notably, Experiment 1 had only six fishing sites and three generalization animals. Therefore, a more complex counterbalancing scheme was used across participants than in the following experiments. Each participant experienced the same four constellations of animal hiding spots, but these were differently labeled (trained site, fully correlated site or anti-correlated site) across participants. Each panel here thus shows the cage drops at a specific constellation of hiding spots but cannot be interpreted in terms of fishing site type.

Fig. S2. Cage drop locations of Experiment 2



Histograms are shown of all individual cage drops in the generalization test rounds across participants. Different colors are used for the different animals. The dots represent the actual hiding spots of the animal. In contrast to Experiment 1, one can for Experiment 2 interpret the label of each fishing site (see Figure 3b). Notice that training was less extensive (in terms of trials and sessions) in Experiment 2 and 3 than in Experiment 1. Histograms were thus much broader. Nevertheless, they do show clear convergence towards the actual hiding spots. Note that this is date from the generalization test rounds in which there was no feedback and inference was needed for appropriate behavior in all other sites than the trained sites (1 and 8).

Fig. S3. Cage drop locations of Experiment 3

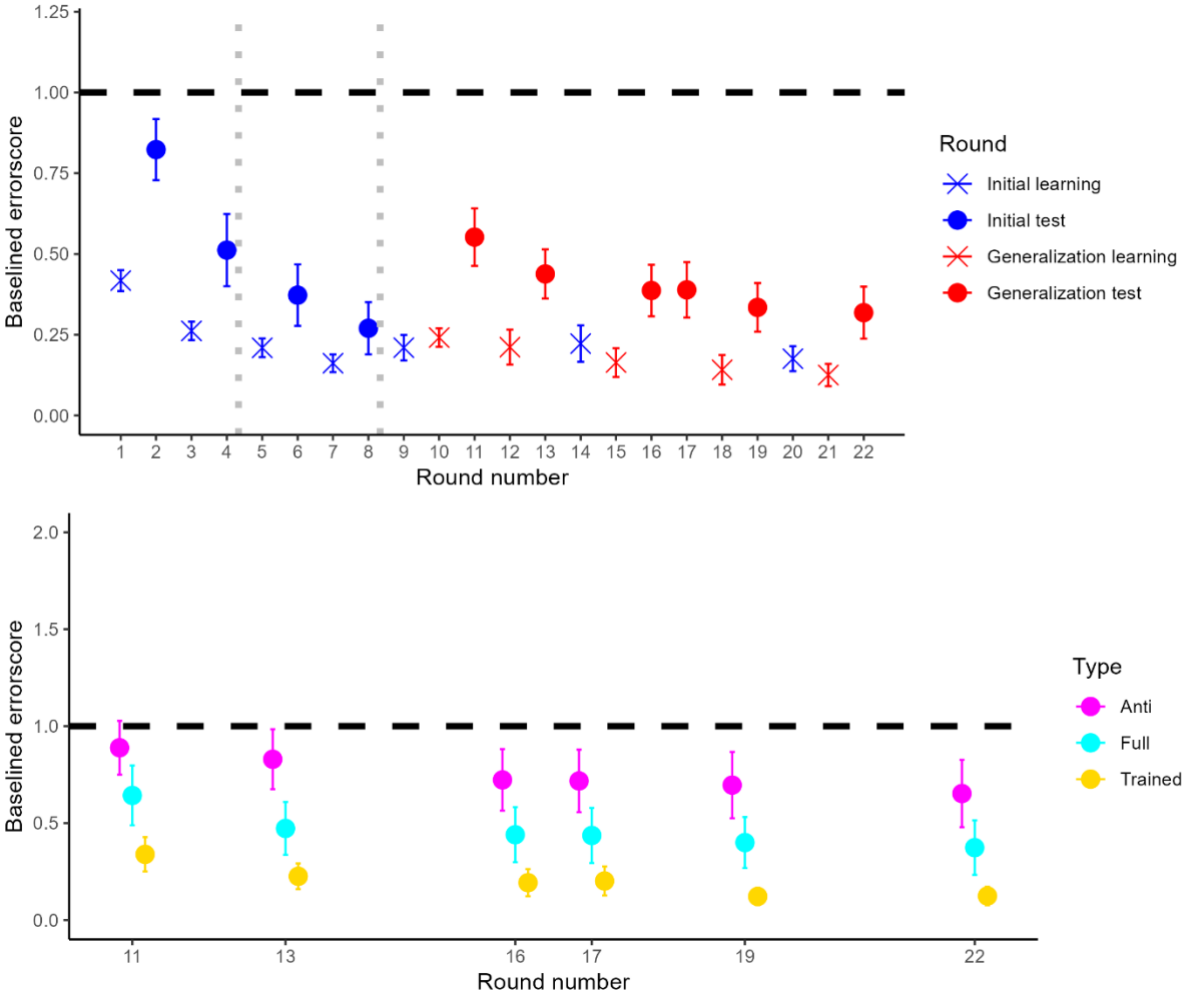


Histograms are shown of all individual cage drops in the generalization test rounds across participants. Different colors are used for the different animals. The dots represent the actual hiding spots of the animal. In contrast to Experiment 1, one can for Experiment 3 interpret the label of each fishing site (see Figure 3b). Notice that training was less extensive (in terms of trials and sessions) in Experiment 2 and 3 than in Experiment 1. Histograms were thus much broader. Nevertheless, they do show clear convergence towards the actual hiding spots. Note that this is date from the generalization test rounds in which there was no feedback and inference was needed for appropriate behavior in all other sites than the trained sites (1 and 8).

Alternative baseline

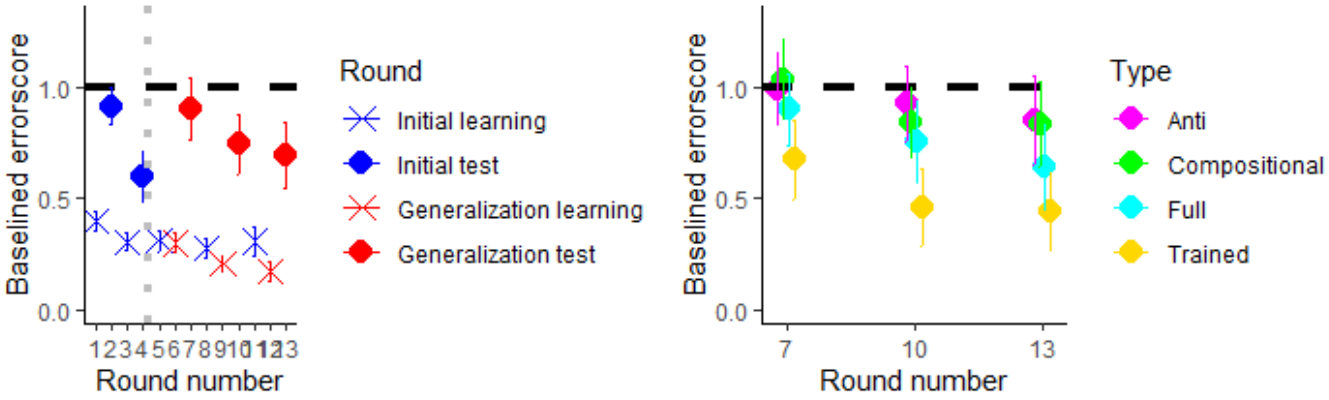
We analyzed and presented our results using a baseline approach. In the main text, a uniform distribution served as the random baseline, comparing participants' behavior to a random agent that could drop the cage at any location with equal probability. Below, we present results using an alternative baseline with a normal distribution, where the random agent could still drop the cage at any location, but with higher probabilities assigned to locations near the center. As described in the main text, we rescaled cage drop locations to fall in between 0 and 1. Correspondingly, we used a normal distribution with mean = .5 and a standard deviation of 0.194. Except for Experiment 2, in which not all results reached significance, the results remain qualitatively similar to those presented in the main text. A standard deviation of 0.194 was chosen to ensure that the baseline distribution covered nearly the entire response range (-1 to 1), which was consistently used by participants. A narrower distribution would not have been representative. This value was selected so that 99% of the distribution fell within the full range, thereby providing a broad yet reasonable baseline.

Fig S4. Randomly distributed baseline results of Experiment 1



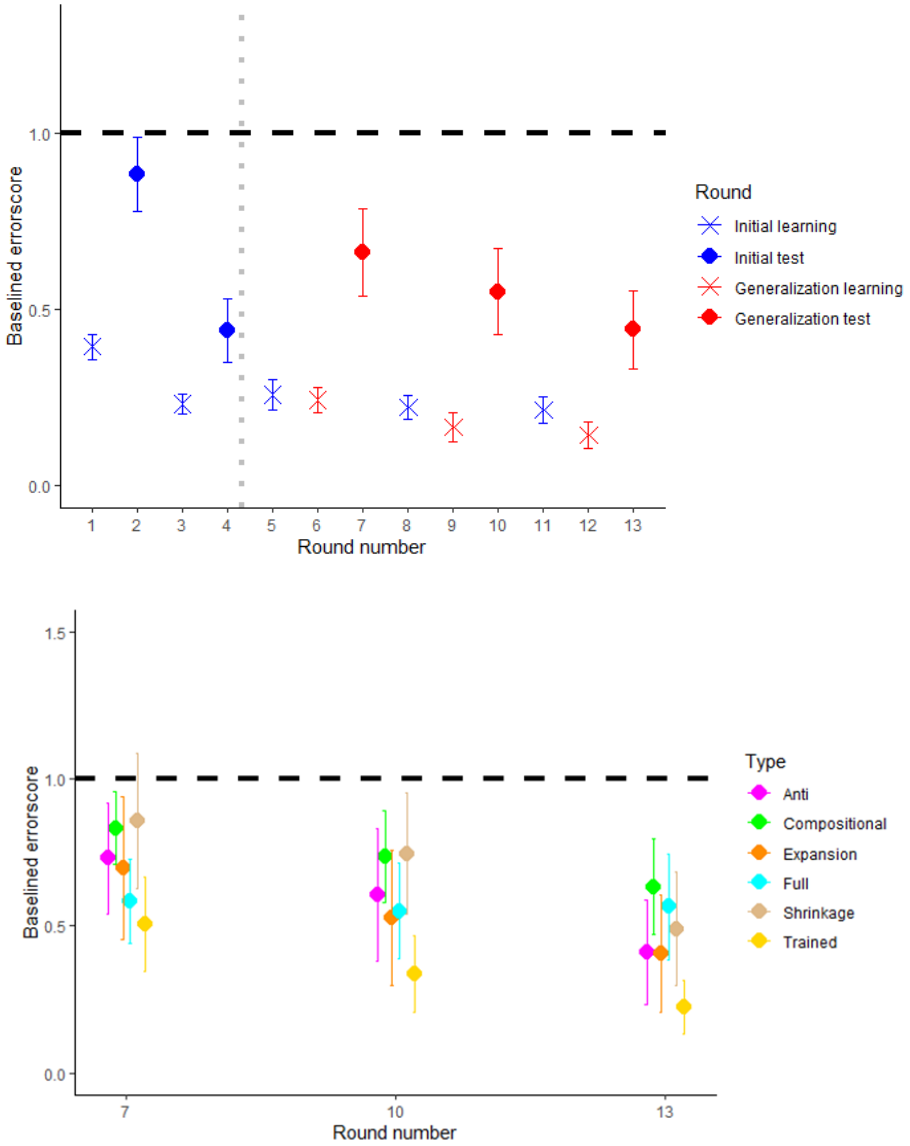
The upper panel illustrates the baselined error score after each round of participant trials. The results indicate that participants exhibit a baselined error score lower than 1 ($t(134) = -14.702$, $p < 0.0001$, $d = -1.27$), suggesting performance better than that of the normally distributed baseline. The lower panel presents the baselined error scores of the specific fishing sites in the Generalization test rounds. In the Generalization test rounds, participants performed significantly better in the "Trained" context compared to the new baseline error score derived from the normal distribution, $t(41) = -41.03$, $p < .0001$, $d = -6.3$. Similarly, performance was also significantly better in the "Full" context, $t(41) = -11.53$, $p < .0001$, $d = 1.78$, and in the "Anti" context, $t(41) = -3.83$, $p < .001$, $d = -0.59$.

Fig S5. Randomly distributed baseline results of Experiment 2



Also in Experiment 2, overall performance was significantly better than the normally distributed baseline (left panel) ($t(123) = -5.66$, $p < 0.0001$, $d = -0.51$). Although the performance in the generalization test rounds (right panel) exhibited a significantly lower baselined error score than 1 in the Trained fishing sites ($t(30) = -6.59$, $p < 0.0001$, $d = -1.18$) and the Fully correlated fishing sites ($t(30) = -3.18$, $p = 0.0017$, $d = -0.57$), this was not true for the anti-correlated ($t(30) = -0.99$, $p = 0.16$, $d = -0.18$) fishing sites and the compositional ($t(30) = -1.36$, $p = 0.093$, $d = -0.24$) fishing sites.

Fig S6. Randomly distributed baseline results of Experiment 3



Also in Experiment 3, overall performance was significantly better than the normally distributed baseline in the test rounds (upper panel) (t(209) = -13.57, p < 0.0001, d = -0.94). Furthermore, even with this normally distributed baseline, performance in the generalization test rounds (lower panel) was significantly better than baseline for all contexts: the trained fishing sites ($t(34) = -12.43$, $p < .0001$, $d = 2.10$), the fully correlated context ($t(34) = -6.30$, $p < .0001$, $d = 1.06$), the anti-correlated context ($t(34) = -6.14$, $p < .0001$, $d = 1.04$), the compositional context ($t(34) = -4.51$, $p < .0001$, $d = 0.76$), the expansion context ($t(34) = -4.50$, $p < .0001$, $d = 0.76$), and the shrinkage context ($t(34) = -3.63$, $p < .001$, $d = 0.61$).