Supplementary document #2

Combined predisposed preferences for colour and biological motion make robust development of social attachment through imprinting (Authors: Momoko Miura, Daisuke Nishi, Toshiya Matsushima)

R functions used for statistic calculations in this study

Levene's test:	leveneTest(x,y), (package "car")
one-sample t-test:	t.test(x, mu=0)
two-sample t-test:	t.test(x,y, var.equal=T,paired=F)
one-way ANOVA:	aov(y~x)
two-way ANOVA:	aov(y~a*b)
two-way ANOVA with repeated measure:	aov(y~a*b+Error(t:a+t:a:b))
Tukey's multiple comparison of means:	TukeyHSD(aov(y~x))
Dunnett's test:	glht(aov(y~x),linfct=mcp(fx="Dunnett")), (package " multcomp")
Steel-Dwass multiple comparison:	Steel.Dwass(x,y), (*)
Pearson's product-moment correlation coefficient:	cor.test(x,y,method="p")
Generalized linear model (GLM) analysis:	glm(y~a + b, family = gaussian)

*: We used the package "RcmdrPlugin.EZR" following Kanda (2013).

Reference

Kanda Y (2013) Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplantation 48:452-458. doi:10.1038/bmt.2012.244

Statistics; number os subjects, sex, and results of one-sample t-test, and Levene's test

In each group, the binary choice data were examined for difference from the chance level by using one-sample t-tests without adjusting p-value by Bonferroni correction. To make comparisons among groups in each experiment, we applied Levene's test to assess the inequality of variances. In the following, results of these calculations are shown in each experiment.

Experiment-1

Four groups were examined; group-1 (n=10; 4 males and 6 females), group-2 (n=10; 4 males and 6 females), group-3 (n=9; 3 males and 6 females), group-4 (n=10; 3 males and 6 females). By our mistake, group-3 included 9 chicks, not 10.

Colour test (Fig. 2, left column):

One-sample t-test for the significant difference from the chance level

Significant positive scores (indicative of yellow preference) appeared in group-1 (a) (t=6.535, df=9, p=0.0001) group-2 (b) (t=3.1987, df=9, p=0.0108), but not in group-3 (c) (t=0.7028, df=8, p=0.4999) and group-4 (d) (t=1.5904, df=9, p=0.1462).

Comparisons among groups

Among the 4 groups, Levene's test failed to detect a significant difference in the variance (F=1.7340, p=0.1779). Parametric test was thus applicable.

Memory test (Fig. 2, right column):

One-sample t-test for the significant difference from the chance level

Significant positive scores (indicative of learned preference) appeared in group-1 (a) (t=6.4714, df=9, p=0.0001), and group-2 (b) (t=4.7803, df=9, p=0.0010), but not in group-3 (c) (t=1.0122, df=8, p=0.3379) and group-4 (d) (t=1.5465, df=9, p=0.1564).

Comparisons among groups

Among the 4 groups, Levene's test failed to detect a significant difference in the variance (F=1.0070, p=0.4012). Parametric test was thus applicable.

Experiment-2

Two groups were examined; group-1 (n=10; 5 males and 5 females), group-2 (n=10; 2 males and 8 females).

Colour test (Fig. 3, left column):

One-sample t-test for the significant difference from the chance level

Significantly positive scores (indicative of yellow preference) appeared in group-1 (a) (t=3.9870, df=9, p=0.0031) and group-2 (b) (t=2.6081, df=9, p=0.0283).

Comparisons between groups

Between the 2 groups, Levene's test failed to detect a significant difference in the variance (F=1.5629, p=0.2272). Parametric test was thus applicable.

Memory test (Fig. 3, right column):

One-sample t-test for the significant difference from the chance level

Significant negative scores (indicative of red preference) appeared in group-1 (a) (t=3.5792, df=9, p=0.0059), but not in group-2 (b) (t=3.7136, df=9, p=0.0048).

Comparisons between groups

Between the 2 groups, Levene's test failed to detect a significant difference in the variance (F=3.7445, p=0.0688). Parametric test was thus applicable.

Experiment-3

Two groups were examined; group-1 (n=10; 5 males and 5 females), group-2 (n=10; 5 males and 5 females).

Colour test (Fig. 4, left column):

One-sample t-test for the significant difference from the chance level

Significantly positive scores (indicative of yellow preference) appeared in group-1 (a) (t=3.7136, df=9, p=0.0048) and group-2 (b) (t=3.5792, df=9, p=0.0059).

Comparisons between groups

Between the 2 groups, Levene's test failed to detect a significant difference in the variance (F= 0.1397, p= 0.7130). Parametric test was thus applicable.

Memory test (Fig. 4, right column):

One-sample t-test for the significant difference from the chance level

Significant negative scores (indicative of red preference) in group-1 (a) (t=4.5187, df=9, p=0.0014), but not in group-2 (b) (t=0.2419, df=9, p=0.8142).

Comparisons between groups

Between the 2 groups, Levene's test failed to detect a significant difference in the variance (F=0.8591, p=0.3662). Parametric test was thus applicable.

Experiment-4

Four groups were examined; group-1 (n=10; 5 males and 5 females), group-2 (n=10; 6 males and 4 females), group-3 (n=10; 5 males and 5 females), group-4 (n=10; 5 males and 5 females).

Colour test (Fig. 6, left column):

One-sample t-test for the significant difference from the chance level

Statistical significance appeared in 3 of these 4; one-sample t-test, group-1 (a) (t=1.0473, df=9, p=0.3223), group-2 (b) (t=3.2778, df=9, p=0.0095), group-3 (c) (t=3.9750, df=9, p=0.0032), group-4 (d) (t=4.0641, df=9, p=0.0028).

Comparisons among groups

Among the 4 groups, Levene's test failed to detect a significant difference in the variance (F=2.5540, p=0.0706). Parametric test was thus applicable.

Memory test (Fig. 6, right column):

One-sample t-test for the significant difference from the chance level

Significant positive scores (indicative of learned preference) appeared in group-1 (a) (t=16.374, df=9, p<0.0001) and group-2 (b) (t=2.7300, df=9, p=0.0232), but not in group-3 (c) (t=0.5962, df=9, p=0.5657) and group-4 (d) (t=0.1186, df=9, p=0.9082).

Comparisons among groups

Among the 4 groups, Levene's test detected a significant difference in the variance (F=3.431, p=0.0270). Parametric test was not applicable.

Experiment-5

Two groups were examined; group-1 (n=11; 5 males and 6 females), group-2 (n=11; 6 males and 5 females).

BM preference test (Fig. 8 top, (a) (b)):

Comparisons among groups

Among the 2 groups tested twice, Levene's test failed to detect a significant difference in the variance (F=0.7132, p=0.5499). Parametric test was thus applicable.

Memory test (Fig. 8 middle, (c) (d)):

Comparisons among groups

Among the 2 groups tested twice, Levene's test failed to detect a significant difference in the variance (F=1.9327, p=0.1398). Parametric test was thus applicable.

Experiment-6

Three groups were examined; group-1 (n=10; 6 males and 4 females), group-2 (n=10; 5 males and 5 females), group-3 (5 males and 5 females).

Comparisons among groups

BM preference test on Day-1

Among the 3 groups tested twice, Levene's test failed to detect a significant difference in the variance (F=1.5356, p=0.2335). Parametric test was thus applicable.

Recent memory test on Day-2

Among the 3 groups tested twice, Levene's test failed to detect a significant difference in the variance (F=0.4423, p=0.6471). Parametric test was thus applicable.

Correlations between Recent memory scores on Day-2 and BM preference scores on Day-1

Because the number of chicks was small (n=10 for each group), correlation coefficient was not supposed to be a highly reliable measure. We therefore avoided to compare the correlation coefficients between the group-1 and -2 after conversion to z-scores. Instead, we constructed Generalized Linear Models (GLMs) for the recent memory score based on the BM preference score and the group as explanatory variables with the interaction term included.

The following tables show the list of 8 models constructed after every possible combinations of terms, and their AICs are indicated (Akaike Information Criteria).

Group-1 and -2

Model	AIC	
#1: recent ~BM	270.7583	
#2: recent ~BM:group	271.9258	
#3: recent ~group	270.3665	
#4: recent ~BM+BM:group	272.7527	
#5: recent ~BM+group	270.0499	(the second-best model)
#6: recent ~BM:group+group	272.1568	
#7: recent ~1 (null model)	271.7081	
#8: recent ~BM+BM:group+group (full model)	268.7647	(the best model)

Coefficients of the best model (#8) were:

BM	1.1151	(p=0.0408)
BM:group	-0.9788	(p=0.0311)
group	259.50	(p=0.1104)

Coefficients of the second-best model (#5) were:

BM	0.3822	(p=0.1670)
group	131.20	(p=0.1350)

The best model (#8) indicates that the BM score contributes to the recent memory score, if the group variable (training on Day-2) was 0 (i.e., group-2); the BM contribution is estimated as 1.1151. If the group variable was 1 (i.e., group-1), the contribution of BM scores is reduced to 1.1151- 0.9788 = 0.1362. The second-best model (#5) indicates that the term of interaction may be disregarded at the cost of reduced reliability of estimates of both coefficients.

<u>Group-1 and -3</u>

Model	AIC	
#1: recent ~bm	281.655	
#2: recent ~bm:group	283.238	
#3: recent ~group	276.3158	(the best model)
#4: recent ~bm+bm:group	283.238	
#5: recent ~bm+group	277.9696	(the second-best model)
#6: recent ~bm:group+group	279.9459	
#7: recent ~1 (null model)	282.3268	
#8: recent ~bm+bm:group+group (full model)	279.9459	

Coefficient of the best model (#3) was:

group	-292.4 (p=0.0080)

Coefficients of the second-best model (#5) were:

BM	0.1666 (p=0.5929)
group	-264.9 (p=0.0302)

The best and the second-best models indicate that group factor (i.e., training on Day-1) contributes to the recent memory score, whereas BM score must be disregarded.

Old memory test on Day-2

Among the 3 groups tested twice, Levene's test failed to detect a significant difference in the variance ($F=1.1817$, $p=0.3221$). Parametric test was thus applicable.