Explanatory Details of Some $\chi^2$ Tests Performed in

Ádám Egri, Miklós Blahó, György Kriska, Róbert Farkas, Mónika Gyurkovszky, Susanne Åkesson, Gábor Horváth (2012)
Polarotactic tabanids find striped patterns with brightness and/or polarization modulation least attractive: An advantage of zebra stripes.
*Journal of Experimental Biology* 215: 736-745

by Ádám Egri and Gábor Horváth

Environmental Optics Laboratory, Department of Biological Physics, Physical Institute, Eotvos University, H-1117 Budapest, Pazmany setány 1, Hungary, gh@arago.elte.hu, adamp39@gmail.com, http://arago.elte.hu

18 March 2014, Budapest

Carl Schwarz was unable to reproduce some $\chi^2$ tests of Egri *et al.* (2012). Here we provide information about the processes how we obtained our results, and we also acknowledge some pitiable typos.

Experiment 1

*Carl Schwarz wrote*: The white-framed black tray without orthogonal white stripes (supplementary material Fig. S1A) captured a significantly higher number ($m = 247$) of tabanids than the white-framed black trays with 2 (Fig. 3A; supplementary material Fig. S1B) and 6 (supplementary material Fig. S1C) orthogonal white stripes ($m = 185$ and 33, respectively) ($\chi^2$ test: $\chi^2 = 108.60$, d.f.$=2$, $P << 0.001$).

I was unable to reproduce this result in my analyses.

**Answer**: Based on the catch numbers in experiment 1 of Egri *et al.* (2012), the $\chi^2$ test could be performed with using the data in Table 1:

<table>
<thead>
<tr>
<th>Number of orthogonal white stripes</th>
<th>Observed by Egri <em>et al.</em> (2012)</th>
<th>Expected by Carl Schwarz</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>247</td>
<td>155</td>
</tr>
<tr>
<td>2</td>
<td>185</td>
<td>155</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
<td>155</td>
</tr>
</tbody>
</table>

**Table 1**: Observed and expected total tabanid catches in experiment 1 when the expected catch numbers are erroneously calculated without taking into consideration the tray-dependent net area of highly and horizontally polarizing black surface parts attracting water-seeking, polarotactic male and female tabanid flies.

In this case the expected catches are $O_{\text{total}}/3 = 155$ for all three trays, where $O_{\text{total}} = 247 + 185 + 33 = 465$. Carl Schwarz calculated the $\chi^2$ tests with equal expected catches for all experiments of ours (Table 1 for experiment 1):
\( \chi^2 \) test for given probabilities

data: fly1.total$tabanids.sum

\[ \chi^2 = 156.4, \text{ df } = 2, \text{ } p < 2.2 \times 10^{-16} \]

(http://people.stat.sfu.ca/~cschwarz/CourseNotes/Reanalysis/ZebraStripes/zebra-anal.html)

Non-metallic smooth (shiny) black horizontal surfaces reflect highly and horizontally polarized light, while white surfaces reflect unpolarized or weakly polarized light. Since the visual attraction of water-seeking male and female polarotactic tabanid flies is elicited by the highly and horizontally polarized light reflected from the horizontal black surface parts of our experimental trays, rather than by the depolarizing white stripes, we modified the evenly distributed expected tabanid catches proportionally to the net black areas of our trays (Table 2). The sound aim of this catch normalization to the net black surface area \( A \) was to show that the decrease of the tabanid catch with increasing number \( N \) of white stripes is due to the increasing \( N \), rather than to the decreasing \( A \).

<table>
<thead>
<tr>
<th>Number of orthogonal white stripes</th>
<th>Observed by Egri et al. (2012)</th>
<th>Expected by Egri et al. (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>247</td>
<td>174.09</td>
</tr>
<tr>
<td>2</td>
<td>185</td>
<td>159.28</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
<td>131.63</td>
</tr>
</tbody>
</table>

**Table 2:** Observed and expected total tabanid catches in experiment 1 when the expected catch numbers are normalized by the net black surface area of the trays.

According to Egri et al. (2012, p. 737), the expected tabanid catches for the different trays should be modified:

Although the whole area of all three trays was the same \( A = 0.25 \text{ m}^2 \), the sum \( A_{\text{black}}(N) \) of their black areas was slightly different owing to the different numbers \( (N) \) of white stripes. The larger the number \( N \) of these stripes, the smaller the highly and horizontally polarizing black area \( A_{\text{black}}(N) \), and thus the smaller the predicted attractiveness to polarotactic tabanids:

\[
A_{\text{black}}(N=0) = (50 \text{ cm} - 2 \times 2 \text{ cm})^2 = 0.2116 \text{ m}^2, \\
A_{\text{black}}(N=2) = 4 \times [(50 \text{ cm} - 3 \times 2 \text{ cm})/2]^2 = 0.1936 \text{ m}^2, \\
A_{\text{black}}(N=6) = 16 \times [(50 \text{ cm} - 5 \times 2 \text{ cm})/2]^2 = 0.1600 \text{ m}^2.
\]

The motivation for making these area calculations is to compare correctly the numbers of trapped tabanid flies: the number \( m \) of tabanids captured by a given tray was divided by the black area \( A_{\text{black}}(N) \), resulting in the surface density \( n = m \times 1 \text{ m}^2/A_{\text{black}}(N) \), giving the number of flies captured by 1 m\(^2\) black tray surface. (see Materials and Methods section in Egri et al. 2012)

Thus, the correctly modified expected tabanid catches are the following:

\[
E(N=0) = O_{\text{total}} \times A_{\text{black}}(N=0)/A = 174.09, \quad (\text{Eqn. 1}) \\
E(N=2) = O_{\text{total}} \times A_{\text{black}}(N=2)/A = 159.28, \quad (\text{Eqn. 2}) \\
E(N=6) = O_{\text{total}} \times A_{\text{black}}(N=6)/A = 131.63, \quad (\text{Eqn. 3})
\]

where \( A = A_{\text{black}}(N=0) + A_{\text{black}}(N=2) + A_{\text{black}}(N=6) = 0.5652 \text{ m}^2 \) is the total black area and \( O_{\text{total}} = 465 \) is the total observed tabanid catch. This surface normalization results in the data in Table 2 from which we performed the \( \chi^2 \) test as follows:
The above are the reason for the difference between the statistical results of Carl Schwarz and Egri et al. (2012).

Note, however, that both approaches have the same statistical outcome: high significance.

**Experiment 2**

**Carl Schwarz wrote:** In experiment 2, the white and black trays filled with salad oil were, respectively, the least and the most attractive to tabanid flies, with 3 versus 145 trapped individuals (supplementary material Tables S3, S4). The smaller the number N of black/white stripes, the more tabanids were trapped; the trays with N = 12, 6 and 2 stripes captured 24, 66 and 138 tabanids, respectively (χ² test: χ² = 88.62, d.f. = 4, P << 0.001; supplementary material Tables S3, S4)

I was unable to reproduce this result in my analyses.

**Answer:** In experiment 2 Egri et al. (2012) the expected tabanid catches were again normalized by the black areas of the trays (Table 3). Here we used a white, three white-striped and a black tray in such a way that the black and white areas of the three striped trays were equal (50-50 %). Thus, the white tray had no black area, and the half of each striped tray was painted to black. The correctly normalized expected catches are:

\[ E(white) = O_{total} \times A_{black}(white)/A = 0, \]  
(Eqn. 4)

\[ E(12 \text{ stripes}) = O_{total} \times A_{black}(12 \text{ stripes})/A = 93.8 \]  
(Eqn. 5)

\[ E(6 \text{ stripes}) = O_{total} \times A_{black}(6 \text{ stripes})/A = 93.8, \]  
(Eqn. 6)

\[ E(2 \text{ stripes}) = O_{total} \times A_{black}(2 \text{ stripes})/A = 93.8, \]  
(Eqn. 7)

\[ E(black) = O_{total} \times A_{black}(black)/A = 94.6, \]  
(Eqn. 8)

where

\[ O_{total} = 3 + 24 + 66 + 138 + 145 = 376, \]

\[ A_{black}(white) = 0, \]

\[ A_{black}(12 \text{ stripes}) = A_{black}(6 \text{ stripes}) = A_{black}(2 \text{ stripes}) = 0.1250 \text{ m}^2, \]

\[ A_{black}(black) = 0.1260 \text{ m}^2, \]

\[ A = A_{black}(white) + A_{black}(12 \text{ stripes}) + A_{black}(2 \text{ stripes}) + A_{black}(black) = 0.501 \text{ m}^2. \]

The software STATISTICA 7.0 used by Egri et al. (2012) treated the white tray’s contribution to the χ²-value as 0.

<table>
<thead>
<tr>
<th>Tray type</th>
<th>Observed by Egri et al. (2012)</th>
<th>Expected by Carl Schwarz</th>
<th>Expected by Egri et al. (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>3</td>
<td>75.2</td>
<td>0</td>
</tr>
<tr>
<td>12 stripes</td>
<td>24</td>
<td>75.2</td>
<td>93.8</td>
</tr>
<tr>
<td>6 stripes</td>
<td>66</td>
<td>75.2</td>
<td>93.8</td>
</tr>
<tr>
<td>2 stripes</td>
<td>138</td>
<td>75.2</td>
<td>93.8</td>
</tr>
<tr>
<td>black</td>
<td>145</td>
<td>75.2</td>
<td>94.6</td>
</tr>
</tbody>
</table>

**Table 3:** Observed and expected numbers of total tabanid catches in experiment 2 when the expected catch numbers are normalized by the net black surface area of the trays.
The difference between the statistical results of Carl Schwarz and Egri et al. (2012) arises again from the lack/presence of normalization of the expected tabanid catches:

<table>
<thead>
<tr>
<th>Carl Schwarz</th>
<th>Egri et al. (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2 = 222.5, \text{df} = 4, p &lt; 2.2 \cdot 10^{-16}$ highly significant</td>
<td>$\chi^2 = 88.6 (107.9), \text{df} = 4, p &lt;&lt; 0.001$ highly significant</td>
</tr>
</tbody>
</table>

We recognized a regrettable typo when comparing the results of Schwarz and Egri et al. (2012): the $\chi^2$-value of 88.6 should be corrected to 107.9 which still means highly significant difference ($p << 0.001$).

**Experiment 3**

Since in experiment 3 of Egri et al. (2012) the ratio of the black and white areas was 1:1 for all striped test surfaces, no surface normalization of the expected tabanid catches was necessary before the $\chi^2$ test. Regarding the test surface with $N = 8$ stripes, Carl Schwarz could not verify the statistical results of Egri et al. (2012):

**Carl Schwarz wrote:** Under all conditions, we found significantly higher numbers of tabanids trapped on the black stripes compared with the white stripes ($\chi^2$ test: 2 stripes, $\chi^2 = 136.08, \text{df} = 1, P < 0.001$; 4 stripes, $\chi^2 = 88.28, \text{df} = 1, P < 0.001$; 8 stripes, $\chi^2 = 5.0, \text{df} = 1, P < 0.025$).\(^12\) (see Results section in Egri et al. 2012)

\(^{12}\)Verified except I was unable to match results for $N = 8$ stripes.

**Answer:**

<table>
<thead>
<tr>
<th>Carl Schwarz</th>
<th>Egri et al. (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2 = 4.263, \text{df} = 1, p = 0.03895$ significant</td>
<td>$\chi^2 = 5.0, \text{df} = 1, p &lt; 0.025$ significant</td>
</tr>
</tbody>
</table>

We acknowledge the recognition of the typo $\chi^2 = 5.0$, which, however, does not concern the final outcome of the statistical significance level, remaining significant.

**Experiment 5**

**Carl Schwarz wrote:** According to the $\chi^2$-test, on surfaces Z2 and Z1, the black stripes trapped significantly more tabanids than the white stripes (supplementary material Table S8).\(^15\) Thus, only when the white and black stripes are wide enough is there a difference in the number of tabanids caught by the black and white stripes.

\(^{15}\)Verified except for results for Z5.

**Answer:** The results of the $\chi^2$ test comparing the tabanid catches of the black and white stripes for the Z5 surface used in experiment 5 of Egri et al. (2012) are the following:

<table>
<thead>
<tr>
<th>Carl Schwarz</th>
<th>Egri et al. (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2 = 0.08, \text{df} = 1, p = 0.7815$ not significant</td>
<td>$\chi^2 = 1.67, \text{df} = 1, p &lt; 0.7$ not significant</td>
</tr>
</tbody>
</table>

We acknowledge the recognition of the typo $\chi^2 = 1.67$ for the Z5 surface, which, however, does not concern the final outcome of the statistical significance level, remaining not significant.
**Summary**

In almost all cases when Carl Schwarz could not reproduce the statistical results of Egri *et al.* (2012), he forgot to normalize the expected catches to the net black area of our striped test surfaces, which is the relevant parameter for the attractiveness to water-seeking polarotactic tabanid flies attracted to highly and horizontally polarized reflected light. On the other hand, we are grateful to Carl Schwarz for finding a few regrettable typos where wrong numerical values appeared in our paper (Egri *et al.* 2012). However, in spite of these typos, our published statistical significance levels are correct in all cases.