Science TECHNICAL RESPONSES

# Response to Comment on “Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions”*

Arnaud Pocheville†[1], Sabine Nöbel[1,2], Guillaume Isabel[3], and Étienne Danchin[1]

[1] *Laboratoire Évolution & Diversité Biologique (UMR 5174); Université Toulouse 3, CNRS, IRD; 118 route de Narbonne, Bat. 4R1, 31062 Toulouse Cedex 9, France.*

[2] *Université Toulouse 1 Capitole, Institute for Advanced Study in Toulouse; 1 esplanade de l’Université, 31080 Toulouse Cedex 6, France.*

[3] *Centre de Recherches sur la Cognition Animale (UMR 5169), Centre de Biologie Intégrative; Université Toulouse 3, CNRS; 118 route de Narbonne, 31062 Toulouse Cedex 9, France.*

† Corresponding author. E-mail: arnaud@pocheville.science. Web: https://arnaud.pocheville.science

**Thornquist and Crickmore claim that systematic experimental error may explain the results of Danchin and colleagues. Their claim rests on mistakes in their analyses for which we provide corrections. We reassert that conformity in fruit flies predicts long-lasting mate-preference traditions.**

Thornquist and Crickmore (2019) argue that from a Bernoulli process of average sample size (63 flies) and average effect (copying probability = 0.68), Danchin et al.’s (2018a; 2018b) data reject (at threshold = 0.05) the null model more often than expected. This is wrong. They counted the same experiment twice (their 89% is consistent with their counting 17, rather than the actual 16, experiments), miscalculated the average sample size (the average is 61.5, not 63), grounded their argument on a non-significant p-value (their 89% $\Leftrightarrow$ $p = 0.11$), and miscalculated their p-value ($p = 0.124$, not their 0.11, taking variation in sample size into account). Their claim is non-significant.

Thornquist and Crickmore argue that significant p-values “cluster more closely towards the usual threshold for accepting a result ($p = 0.05$) … than expected … ($p = 0.064$)”. This is wrong. In addition to the mistakes above which percolated into their Kolmogorov-Smirnov (KS) test, they wrongly compared Danchin et al.’s p-value distribution to a distribution computed for an experiment of average sample size, while p-value distributions precisely depend on sample sizes (which vary in this dataset). Our own computation of 10,000 two-sided KS-tests performed on p-value distributions obtained on simulated sets of 16 experiments of sample sizes equal to ours yielded ca. 9,500 tests with a p-value > 0.05 and an average p-value = 0.43 (figures 1, 2, 3). Thornquist and Crickmore’s reported clustering of p-values is non-significant.

Thornquist and Crickmore argue that “[non-significant] p-values [in the GLMM] are skewed strongly towards 1, indicating far lower variance in preference than would be expected from an unbiased set of experiments”. This is wrong. We had long noticed that variance in mate-copying was low compared to an average Bernoulli process. This is however exactly what is expected under a mixture of Bernoulli processes (figure 4). Several factors may affect mate-copying, that will lead to a reduced variance among treatments that are performed in parallel. This is a well-known mathematical feature of good experimental design. The variance will be even lower if females show anticonformity when stressed and a very high capacity to learn in the best conditions.

Thornquist and Crickmore argue that transmission chain results more closely resemble the predicted mean of a Bernoulli process than expected by chance. This is wrong. Their argument is grounded on a non-significant p-value (their $p = 0.07$, from their 10,000 simulations), which we were again unable to reproduce. Our own simulations yielded $p = 0.56$; that is, 44% of simulated transmission chains were closer to the mean trajectory than the observed trajectory (1,000,000 simulations, figure 5). This is to be compared to $p < 10^{-6}$ for the null hypothesis: all chains simulated assuming no copying were closer to the mean null trajectory than the observed trajectory. More generally, Thornquist and Crickmore’s discarding results because data match the model’s predictions is a reasonably problematic scientific procedure (figure 6).

Thornquist and Crickmore speculate that male coloring may be one source of experimental bias, the quantity of powder possibly affecting demonstration and

---

*Arnaud Pocheville et al. (2019). “Response to Comment on “Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions””. In: *Science* 366.6462, p. 194. DOI: 10.1126/science.aaw9549. Version accepted for publication (3 June 2019) before editorial revisions. Copy-edited, 29 December 2020.

test males of the same color. This is wrong. When females freely chose the males, demonstration and test males did not come from the same coloring vial. When females were introduced as already copulating with a male of a given color, this male was randomly imposed by the experimenter (Dagaeff et al. 2016). Their speculation about a “functional criterion for knowing whether enough color remains [being] mate-copying [itself]” is incoherent, since coloring is finished before mate-copying is assessed.

Thornquist and Crickmore note that Danchin et al. reported a significant effect of experimenter identity on mate-copying. Danchin et al. also reported, importantly, that controlling for it did not change the results.

Thornquist and Crickmore propose a model of transmission chain based on Danchin et al.’s. Their assumption that flies remain active at low air-pressure is surprising. Their model (their fig. 2B) fits neither Dagaeff et al. (ibid.) nor Danchin et al. (2018a; 2018b) data, where the inflexion point occurs around 1007 hPa for short demonstration protocols and air pressure does not correlate with mate-copying in long demonstration protocols. This invalidates their analysis.

To conclude, we thank Thornquist and Crickmore for their comment and reiterate that we welcome replications of this study with excitement. We warmly invite researchers interested in studying social learning in Drosophila to visit us, as others have already done, to learn all the tricks of the trade and develop together new approaches for invertebrate models of animal culture.

## Acknowledgements

We thank Marlène Framit, Aliénor Jeliazkov, Maël Montévil and Aurélien Pocheville for discussions and help with figures, tests and simulations.

## References

Dagaeff, Anne-Cécile et al. (2016). “Drosophila mate copying correlates with atmospheric pressure in a speed learning situation”. In: *Animal Behaviour* 121, pp. 163–174. DOI: 10.1016/j.anbehav.2016.08.022.

Danchin, Etienne et al. (2018a). “Data from: Cultural flies: conformist social learning in fruit flies predicts long-lasting mate-choice traditions”. In: *Dryad Digital Repository*. DOI: https://doi.org/10.5061/dryad.683870h.

Danchin, Étienne et al. (2018b). “Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions”. In: *Science* 362.6418, pp. 1025–1030. DOI: 10.1126/science.aat1590.

Thornquist, Stephen and Michael A. Crickmore (2019). “Comment on “Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions””. In: *Science* 366.6462, p. 194. DOI: 10.1126/science.aaw8012.

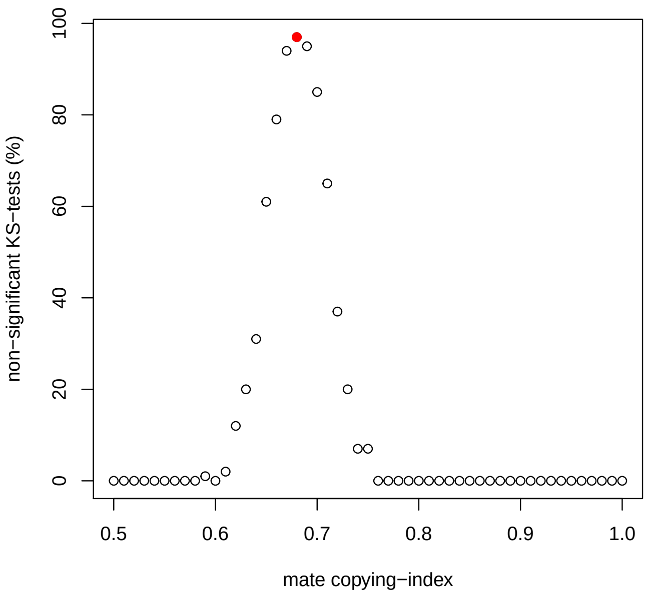

Figure 1: Percentage of non-significant Kolmogorov-Smirnov tests between Danchin et al.’s 16 p-values and distributions simulated for an average Bernoulli process depending on mate-copying index (MCI). The red dot corresponds to the observed average MCI = 0.68 (10,000 simulations).

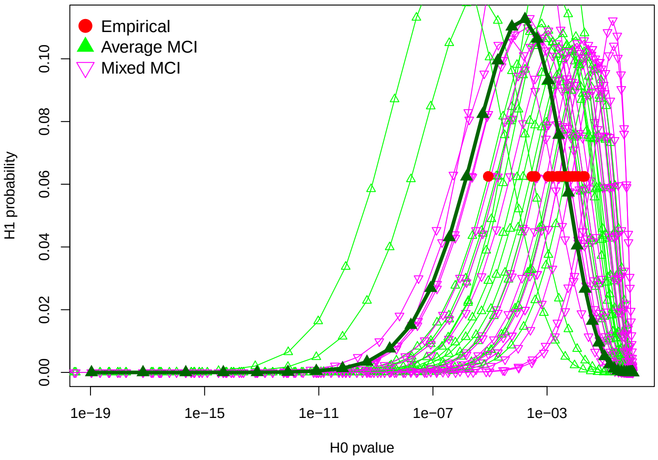

Figure 2: Probability distributions for p-values under H0 if H1 is true. Dark solid line: exact values for an average Bernoulli process similar to Thornquist and Crickmore’s (N = 62, p = 0.68). Red dots: empirical distribution. The empirical distribution looks offset due to the differential impact of some experiments on the average mate-copying index (MCI) and the distribution of p-values (an heterogeneity obscured by Thornquist and Crickmore’s averaging). Green and pink lines: probability distributions computed for two sets of 16 digital experiments originating from an average process (MCI = 0.68, green) or a mixture of Bernoulli processes (MCI between 0.5 and 1 with an average of 0.68, pink).

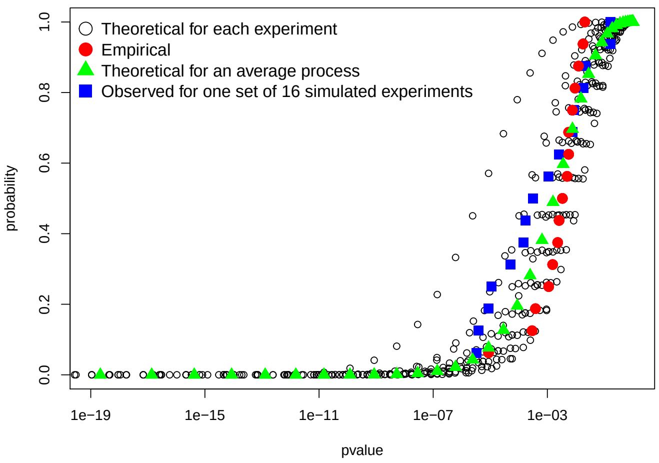

Figure 3: Cumulative distribution functions for the empirical set of 16 p-values (red dots), for Thornquist and Crickmore's theoretical average process (green triangles), and for one set of 16 simulated experiments (blue squares). In the background are the exact cumulative distributions for each experiment (empty dots). The variation comes from their different sample sizes and effect.

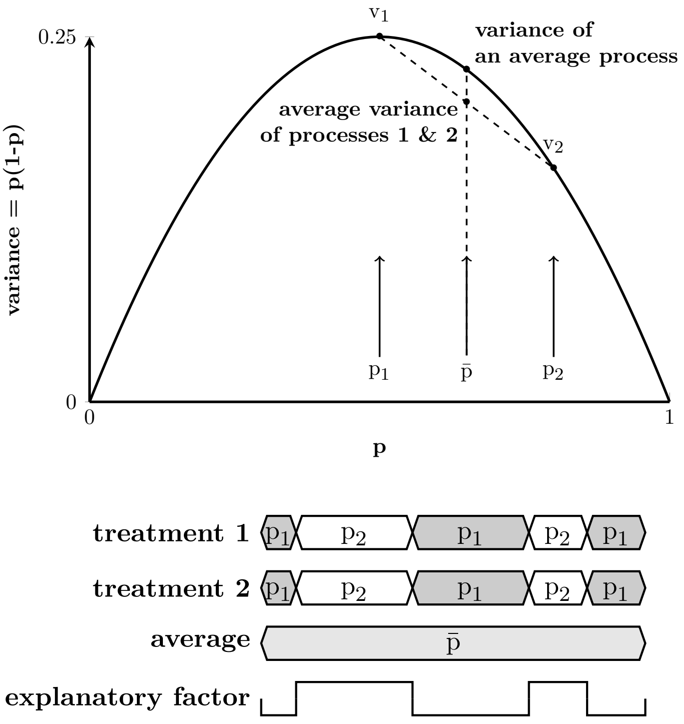

Figure 4: Variance will be lower under a mixture of Bernoulli processes than under an average Bernoulli process. Performing treatements in parallel instantiates this mathematical fact.

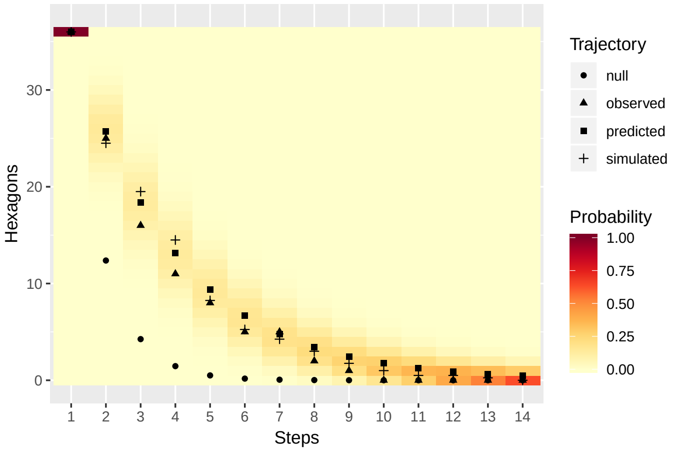

Figure 5: Transmission-chain model in Danchin et al. (2018b). Colors: exact probabilities. Symbols: trajectories illustrated in the original paper. Contra Thornquist and Crickmore (gray area in their figure 1C), the density is not uniform, but highly skewed towards the mean.

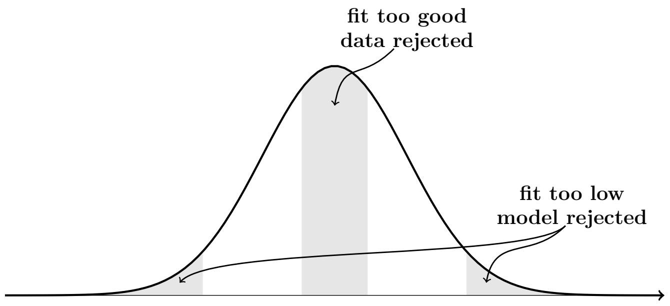

Figure 6: Thornquist and Crickmore's method of discarding results when data fit predictions doubles the rate of wrong rejection of correct models. Asymptotically, discarding results because data match or mismatch any moment of a predicted distribution raises the rejection rate to 100% of correct models.

27 February 2019; accepted 14 June 2019

10.1126/science.aaw9549