No support for honest signalling of male quality in zebra finch song

Martin Bulla and Wolfgang Forstmeier

Alam et al.[1] claim to have discovered a song feature, called "path length", that honestly signals male fitness and is therefore preferred by all females. We see no statistical support for this claim in the original data. (1) The main finding about path length being an honest signal of quality (Fig. 4c) results from a statistical artefact, the regression of *y minus x* over *x*, which creates an illusory effect where none exists. (2) The low technical repeatability of path length measure further questions its use as a proxy of male quality. Consequently, the conclusion that females generally prefer songs with long path lengths in playback experiments is unconvincing and also inconsistent with general knowledge about mate choice in the studied species[2-6].

(1) Alam et al.[1] introduce a new song feature, "the spread of song in latent space, defined by the minimum path length connecting song syllables", and hypothesise that "it should be more difficult [for juvenile males] to imitate [, i.e. learn,] long-path-length songs". However, they found that "all pupils learned reasonably well… and there was no correlation between adult similarity [i.e. the acoustic similarity of pupil's song to tutor song] and the path length of the tutor's song (Fig. 4b)." However, Alam et al.[1] do not stop there, claiming that "comparing the change in path length of the pupil's song with the tutor's song… a significant negative correlation [emerges] (Fig. 4c)" and conclude that "[j]uvenile birds tutored by birds with short-path-length songs were able to match or exceed the path length of their tutor, whereas birds tutored by birds with long-path-length songs struggled to match those path lengths by adulthood". This result is due to a statistical artefact of regressing *y minus x* over *x*.

To illustrate the case (top panels in Fig. 1), we randomly generated 1000 values of *x* and, independently, 1000 values of *y*. Accordingly, *x* and *y* are uncorrelated (top left panel). However, when the difference (*y*-*x*) is regressed on *x*, a strong negative relationship emerges (r = -0.7; top middle panel). Similarly, and for completeness, *adding y and x* and plotting it over *x* yields a strong positive relationship (r = 0.7; top right panel). The negative and positive correlations arise simply because x is included in both axes.

The same applies to the Alam et al.[1] data behind Fig. 4c, where the negative relationship between *pupil minus tutor path length* and *tutor path length* arises because the tutor path length is included on both axes (bottom panels in Fig. 1 below). The absence of any (also non-linear) relationship between pupil and tutor path length (bottom left panel) highlights that there is no evidence in the data that long song paths are difficult to learn, and thus an honest signal of male quality. In the case of Alam et al.[1], either pupils do not learn a path length from their tutor, or Alam et al.'s path length is not a biologically meaningful song parameter to begin with. Either of these interpretations is valuable to the song research community.

Importantly, demonstrating that path length is an indicator of individual quality requires that path length positively correlates with a meaningful measure of individual phenotypic quality based on strong proxies for fitness (e.g. [3]). Although isolated studies have repeatedly reported some associations between aspects of zebra finch song and male quality (reviewed by [7]), none of these associations seemed to hold up in follow-up studies[7-10]. We have seen the same repeatedly in our laboratory, where initial findings never held up when examined with >600 males (>200 song parameters, including machine-learning approaches; [3,11,12]). As a result, zebra finch song has repeatedly been seen as an identity signal, i.e., a name-tag (short, individually distinct, and unspectacular; [11,13,14]), rather than a signal of quality. If name tags were to be honest indicators of quality, then females should at least agree on which name tag is attractive. This is not the case[6 and as we discuss next].

(2) To test whether females prefer songs with longer path lengths, Alam et al.[1] used only three independent pairs of song stimuli (short vs long path length) in their choice chamber experiments. As these six song stimuli were artificially created, it is possible that some of them happened to sound either interesting or aversive to the ears of female zebra finches, leading to either curiosity or avoidance. To disentangle the desired effects from chance, more song pairs are needed, as the probability that all three long-path songs are approached by females by chance alone is quite high ($0.5^3 = 12.5\%$; for details see[15,16]).

Indeed, Alam et al.'s[1] Fig. 3d gives the impression of unanimous female choice for long-path songs. Such a finding contradicts what has been known for more than twenty years[6 and references therein], namely that female zebra finches rarely agree on which male is attractive[2-5]. If there is no repeatability of male attractiveness across different females then there cannot be a magic "X-factor"[17] that makes some males more attractive than others. Interestingly, we recently found that even the individual-specific female preferences for particular males are independent of their song[5]. Furthermore, Alam et al.[1] Fig. 3d is probably only valid for the path lengths from a given latent space (called UMAP), since path lengths of a specific individual, generated from multiple latent spaces, have very low technical repeatability (P=0.06, R^2=0.242; [1]). To have a meaningful stimulus pair, given the low repeatability, one needs to contrast a song with a very short path length against a song with a very long path length (see pair 2 in Fig. 2 below). If the difference between the two songs in a pair (see pair 3 and perhaps even pair 1) is too small, a new latent space simulation is likely to reverse the assignment of which of the two songs is considered to have the longer path. Alam et al.'s[1] tests of pairs 1 and 3 are therefore largely uninformative. Moreover, in the only informative pair 2, the long-path song was not preferred by two out of three females (Fig. 2). It is unclear why the most contrasting song pair (pair 2) was used the least and why more song pairs with highly contrasting path lengths were not created and tested. With 13 test females available, one could have designed 13 pairs of stimulus songs, with each providing a strong contrast between a very short and a very long song path. Such a design would have been much more informative than using a single contrasting song pair (pair 2) tested once on only 3 females.

To conclude, there have been many attempts over the past few decades to identify a song feature that makes zebra finch songs generally attractive[7]. To the best of our knowledge, all these attempts have failed, reflecting the lack of major variation in male attractiveness[2-6]. Alam et al.'s[1] path length appears to be no exception.

References

1 Alam *et al.* The hidden fitness of the male zebra finch courtship song. *Nature* **628**, 117-21 (2024). https://doi.org/10.1038/s41586-024-07207-4

2 Ihle *et al.* Fitness Benefits of Mate Choice for Compatibility in a Socially Monogamous Species. *PLoS Biol* **13**, e1002248 (2015). https://doi.org/10.1371/journal.pbio.1002248

3 Wang *et al.* No mutual mate choice for quality in zebra finches: Time to question a widely-held assumption. *Evolution* **71**, 2661–76 (2017). https://doi.org/10.1111/evo.13341

4 Wang *et al.* The role of genetic constraints and social environment in explaining female extra-pair mating. *Evolution* **74**, 544-58 (2019). https://doi.org/10.1111/evo.13905

5 Wang *et al.* Is female mate choice repeatable across males with nearly identical songs? *Anim Behav* **181**, 137-49 (2021). https://doi.org/10.1016/j.anbehav.2021.09.001

6 Forstmeier & Birkhead Repeatability of mate choice in the zebra finch: consistency within and between females. *Anim Behav* **68**, 1017-28 (2004). https://doi.org/10.1016/j.anbehav.2004.02.007

7 Riebel. Vol. 40 *Advances in the Study of Behavior* Ch. 6, 197-238 (2009). https://doi.org/10.1016/s0065-3454(09)40006-8.

8 Gil *et al.* Early condition, song learning, and the volume of song brain nuclei in the zebra finch (Taeniopygia guttata). *J Neurobiol* **66**, 1602-12 (2006). https://doi.org/10.1002/neu.20312

9 Templeton *et al.* Does song complexity correlate with problem-solving performance in flocks of zebra finches? *Anim Behav* **92**, 63-71 (2014). https://doi.org/10.1016/j.anbehav.2014.03.019

10 Kriengwatana *et al.* Auditory discrimination learning in zebra finches: effects of sex, early life conditions and stimulus characteristics. *Anim Behav* **116**, 99-112 (2016). https://doi.org/10.1016/j.anbehav.2016.03.028

11 Forstmeier *et al.* The genetic basis of zebra finch vocalizations. *Evolution* **63**, 2114-30 (2009). https://doi.org/10.1111/j.1558-5646.2009.00688.x

12 Wang *et al.* Machine learning reveals cryptic dialects that explain mate choice in a songbird. *Nat Commun* **13**, 1630 (2022). https://doi.org/10.1038/s41467-022-28881-w

13 Tibbetts & Dale Individual recognition: it is good to be different. *Trends Ecol Evol* **22**, 529-37 (2007). https://doi.org/10.1016/j.tree.2007.09.001

14 Miller The acoustic basis of mate recognition by female Zebra finches (Taeniopygia guttata). *Animal Behavior* **27**, 376-80 (1979). https://doi.org/10.1016/0003-3472(79)90172-6

15 Kroodsma *et al.* Pseudoreplication in playback experiments, revisited a decade later. *Anim Behav* **61**, 1029-33 (2001). https://doi.org/10.1006/anbe.2000.1676

16 Kroodsma Using appropriate experimental designs for intended hypotheses in 'song' playbacks, with examples for testing effects of song repertoire sizes. *Anim Behav* **40**, 1138-50 (1990). https://doi.org/10.1016/S0003-3472(05)80180-0

17 Howe & Thompson AI hears hidden X factor in zebra finch love songs. *Nature*, (2024). https://doi.org/10.1038/d41586-024-00864-5

Code availability

Code to generate the figures is available at https://github.com/MartinBulla/rebuttal_alam_2024.

Acknowledgements

We are grateful to all colleagues that invested into discussing the matters with us. M.B. was supported by the Research Excellence in Environmental Sciences Project (REES 003) from the Faculty of Environmental Sciences, Czech University of Life Sciences Prague.

Authors and Affiliations

Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Prague 6 – Suchdol, Czech Republic

Martin Bulla (bulla.mar@gmail.com)

Max Planck Institute for Biological Intelligence, Eberhard-Gwinner-Str., 82319 Seewiesen, Germany

Wolfgang Forstmeier (wolfgang.forstmeier@bi.mpg.de)

Competing interests

The authors declare no competing interests.

Contributions

M.B. and W.F. contributed equally.

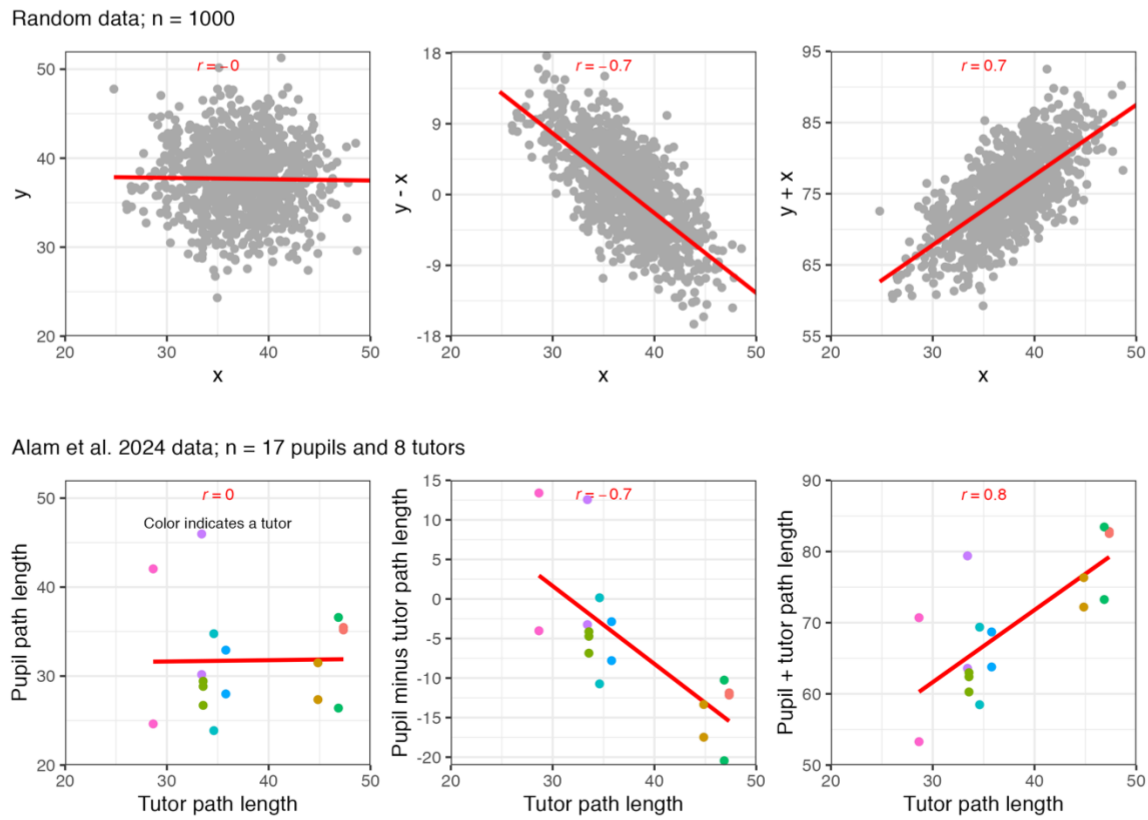

Figure 1 | Illustration of illusory relationships when an x-variable is included in both axes. Top panels are based on the 1000 randomly sampled values of x and y, **bottom** panels depict data from Fig. 4c of Alam et al. 2024. In **both**, the top and bottom panels, dots represent individual data points, their colour in the bottom panels highlights individual tutors. Red lines represent ordinary least-square regressions. *r* is a Pearson's correlation coefficient. **Left** panels highlight no relationships in the data. **Middle** panels highlight negative relationships, and **right** panels positive relationships, both arising from including an x-variable also in the y-axis.

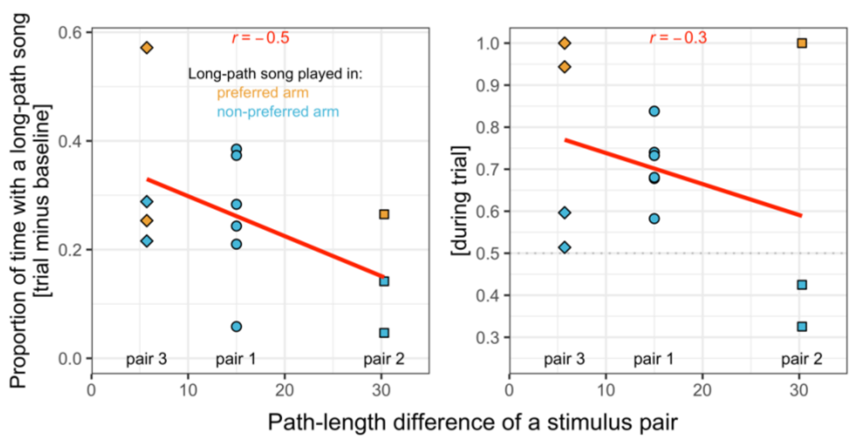

Figure 2 | Female preferences as a function of how different the two song stimuli are. The x-axis shows the difference in path lengths between the two songs of a stimulus pair. The y-axis shows the change in the proportion of time spent in the choice-chamber arm with the long-path playback during the trial compared to baseline, i.e. the mean of pre- and post-trial values (**left**), or the proportion of choice time (excluding the neutral arm) spent in the arm with the long-path playback (**right**; values measured from green versus black bars during playback in the Extended Data Fig. 6). The shapes represent individual observations for each stimulus pair, the dot colour indicates whether a long-path song was played in the arm preferred by the female during the pre-trial period (orange) or in the other arm (blue). Red lines represent ordinary least-square regressions. *r* is a Pearson's correlation coefficient. Note, if path length was the underlying cause of female preferences, a positive regression slope (not a negative one) would be expected. Unexpectedly, the most extreme pair 2 shows the weakest (not the strongest) effects, and pair 1 shows the clearest effect. It is hence unclear what was special about pair 1, apart from the path length difference, which was rather modest. In addition, the path length difference in pair 3 is so small that minor changes in UMAP space could reverse the assignment of which stimulus in pair 3 actually represents the longer path.