A reply to Srinivasan et al.'s rebuttals

Srinivasan has provided three responses to our manuscript and independent comments that have been raised by others. You can find his response here:

1) https://pubpeer.com/publications/61066107845FC9C9D33B257943E32B (by MV Srinivasan and J Tautz)
2) https://arxiv.org/pdf/2408.11520 (by MV Srinivasan, J Tautz, and GW Stuart)
3) https://arxiv.org/pdf/2408.07713 (by GW Stuart)

Srinivasan's coauthor on two of these replies is GW Stuart, an honorary member of the Melbourne School of Psychological Sciences:
https://findanexpert.unimelb.edu.au/profile/3242-geoff-stuart

1)
Let's begin with the response to the PubPeer comment by MV Srinivasan and J Tautz. The PubPeer comment (not authored by us) noted that in Figure 2, the means and SEMs of three separate experimental groups are identical. We are reproducing the figure included by the anonymous author here (left panel):
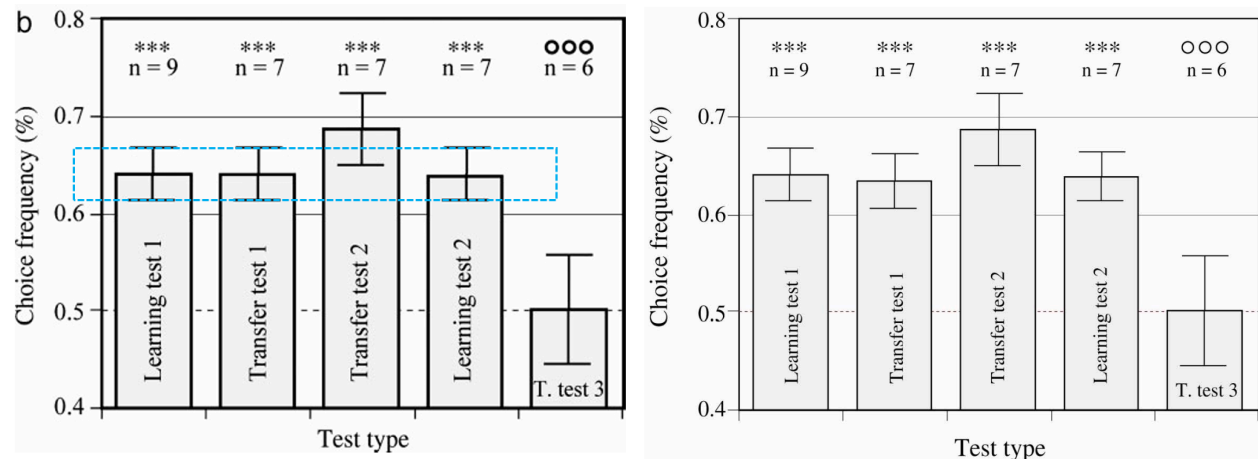
In Srinivasan and Tautz's response, they allege that while the mean and SEM values are similar, they are not identical. According to them, the values appear identical in the published figure (left panel above) because "the resolution of the image published by PNAS is inadequate." Moreover, they provide "a copy of the figure submitted to PNAS" (right panel above).

Upon close inspection, it immediately becomes clear that **the figures are not the same**. The font size is larger in the figure published by PNAS (e.g. see the space between "T. test 3" and the edge of the bar), and 2-3 dashes between the bars become 5 dashes. These changes are not caused by a change in resolution, and unless PNAS decided to change the space between dashes in the figure, **it raises the question of where this new figure came from. Especially because, according to Srinivasan and Tautz themselves, "the relevant datasets for the figure are no longer available."**

2)

Srinivasan, Tautz, and Stuart wrote a point-by-point response to our manuscript that they published on the arXiv. Unfortunately, though perhaps unsurprisingly (given Srinivasan's previous responses), the responses are filled with personal insults and disingenuous strawman arguments. We highlight a few examples below, but it is important to note that all clear instances of data duplication are dismissed as "minor typographical errors" and all other concerns we raised are dismissed due to our field of research.

Already in the third paragraph, Srinivasan claims, "The authors have also falsely claimed that inadvertent data entry errors in two papers, which have since been corrected and published in The Journal of Experimental Biology, are examples of 'data manipulation and duplication'." The occurrence of data duplication is a fact that both Srinivasan and The Journal of Experimental Biology (JEB) admitted in what they constitute as "corrections" (available here and here). We are copying below what Srinivasan repeatedly calls a "correction" in his response (and which apparently is good enough for JEB):

"The graph in Fig. 3A labelled 'still air' with N=19 is reused in Fig. 3B with N=35 and labelled '14 cm tunnel control'. The control graph in Fig. 7 is labelled with a tunnel length of 7.6 m and N=71, which is different to the identical graph shown in Figs 2, 3 and 5 [continued listing of the occurrence of duplicated data for different experiments]. Unfortunately, the **original data, analyses and figure plots are not available to determine the reasons for the discrepancies**. [...] The 1996 paper is *likely* to contain the correct values [...]. Srinivasan takes sole responsibility for the errors and apologises for any inconvenience caused, but **states that the issues do not alter the overall results and conclusions of the paper [that were concluded based on the data in question].**"

According to Srinivasan et al. (and JEB) simply guessing at what the correct data looked like (or admitting that it does not exist) constitutes a valid "correction" for instances of identical data being reported for different experimental conditions:
- "it is very **likely** that the width of the narrow tunnel was not 7 cm (typographical error), but 11cm"
- "**the best guess** would be that the number n=88 given in the earlier publication (1996) is the correct figure, and the number n=56 provided for the same experiment in the subsequent publication (1997) is an error of numerical entry."
- "Srinivasan does not have the original data, but it appears that Fig. 7 does not contain the correct graph for the searching distribution with the landmark positioned at Unit 9, which was a separate experiment from those described in Figs 2, 3 and 5."
- "It is impossible, at this stage, to specify which sample size is the correct one (N=19 or N=35)."

Next, Srinivasan et al. falsely claim that "L&P have not understood our objective in calibrating the odometer" and "L&P are not aware of these nuances, because they do not work in this field" because we supposedly did not understand that "[Srinivasan et al.] deliberately excluded

consideration of the intercept in the waggle duration." This is false, and Srinivasan et al. themselves cite the relevant sentence from our manuscript, where we wrote: "We noticed that the data does not support that 186 m of outdoor flight is encoded by a waggle duration of 350 ms *unless the y-intercept (96 ms) is ignored*."

Next, Srinivasan et al. attempt to explain why the reported data does not match the reported methods. To recap, in the 2000 Science paper by Srinivasan et al. the authors report the following method: "Waggle dances were video-filmed at 25 frames per second and were later played back for frame-by-frame analysis. The duration of each waggle phase was measured in terms of the number of frames over which it occurred." Since 25 frames per second constitute frames of 40 ms each, the resulting waggle durations should be multitudes of 40. We showed that this is not the case for the reported means in combination with the reported number of replicates.

Srinivasan et al. claim that the means do not match the reported methods because they are "rounded values (rounded to one decimal place)." Since this still does not explain the discrepancy for 4 out of 10 values, the authors go on to explain that "**these mismatches are likely due to minor inconsistencies in the rounding operations (rounding up instead of down or vice-versa)**, or dropping one or more trials post video analysis" (which should still result in multitudes of 40).

One may wonder at this point, who would accept these explanations as "corrections"? **We are deeply disappointed to report that Science has joined JEB in giving a pass to this incredibly low standard of scientific integrity.** Even if no manipulation of data has occurred, the fact remains that we do not have the data that belongs to the reported experiments (this is certainly true for the data published in JEB as admitted by both JEB and Srinivasan), and without coherent explanations for the discrepancies, we noted it seems the papers ought to be retracted.
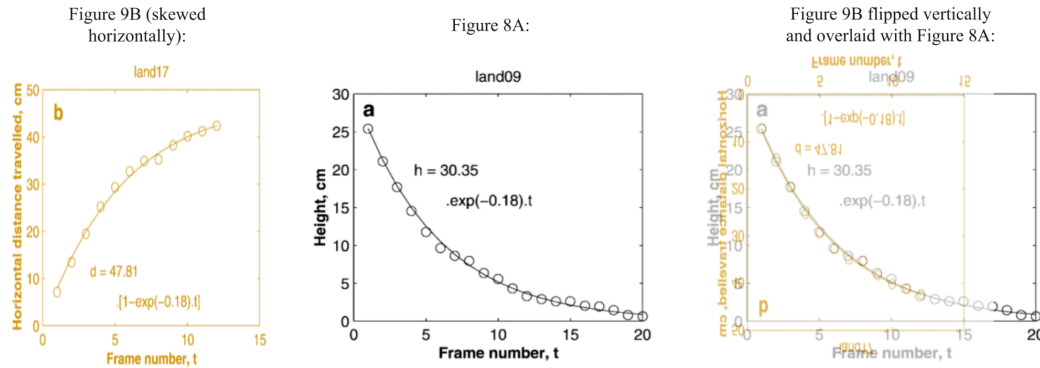
The authors end this section with an insulting strawman argument: "L&P's computation of the first and second set of numbers in the rightmost column of their Table S1 is not logical: there is no expectation for these numbers to be integers. The video camera was not used to count the number of bees, the number of dances, or the number of waggle phases. These numbers were obtained from manual observation of the recorded video dances; they have nothing to do with the frame rate of the video, which is irrelevant in this context." We never claim this; in the third column of our Table S1, we are simply reproducing the different numbers of replicates reported in the original paper by Srinivasan et al. Since the original paper does not specify which n was used to compute the mean, we show that none of them match the reported frame rate.

Regarding the highly similar data points at the end or beginning of datasets that we report in our Supp. Fig. 2, **Srinivasan et al. simply conclude, "The close similarity of some data points (or disparity in others) is irrelevant"** since the axes do not match and "these are all distinct data sets." Given the numerous instances of Srinivasan et al. reporting identical data for reportedly separate datasets, it is difficult to take this explanation seriously.

Moreover, **according to Srinivasan et al., our claim that the data points shown in panel A below are highly similar "is extraordinary":**

A

Figures from Srinivasan, M., Zhang, S., Chahl, J. *et al.* How honeybees make grazing landings on flat surfaces. *Biol Cybern* 83, 171–183 (2000). https://doi.org/10.1007/s004220000162

Seemingly manipulated, highly similar data reported for different experiments

Figure 9B (skewed horizontally):

Figure 8A:

Figure 9B flipped vertically and overlaid with Figure 8A:

B

Seemingly manipulated, identical data reported for different experiments across different papers

Figure 5A (skewed vertically) from M. V. Srinivasan, S. W. Zhang, M. Lehrer, T. S. Collett; Honeybee Navigation *En Route* to the Goal: Visual Flight Control and Odometry. *J Exp Biol* (1996). doi: https://doi.org/10.1242/jeb.199.1.237

Figure 3B from Srinivasan, M., Zhang, S., Chahl, J. *et al.* How honeybees make grazing landings on flat surfaces. *Biol Cybern* 83, 171–183 (2000). https://doi.org/10.1007/s004220000162

Overlay of Figure 3B (from 2000) and Figure 5A (from 1996):
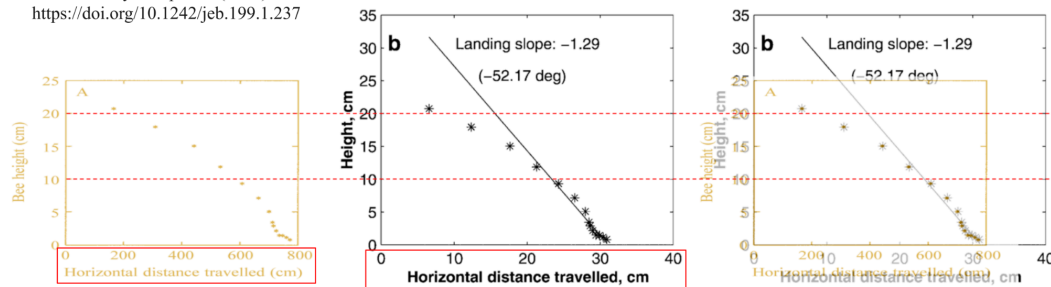
Regarding panel B, the authors explain, "The reason for this discrepancy [in the x axis] is **most likely** that the horizontal distance travelled was erroneously multiplied by the video frame (25 frames/s) in labelling the X-axis labelling." Again, the authors assert that "this error does not affect any of the analyses, results, or conclusions." and later repeatedly dismiss this error as "we have already shown above that there are no problems with the data in the 2000 [Biological Cybernetics] paper."**

Lastly, here are two notes regarding Srinivasan et al's responses to our Table 1:
1. Srinivasan et al. state that "We see no problem with the re-use of data from certain control experiments conducted in the 1996 JEB study in the 1997 JEB study: At the end of the Introduction section of the 1997 paper we have stated explicitly that **"A preliminary account of this work is given in Srinivasan et al. (1996)." This sentence apparently should inform readers that the different experiments described in the 1996 and 1997 papers did not have their own controls.**

2. Srinivasan et al. say, "It is claimed that, in the 2004 paper on Landing Strategies in Honeybees and Applications to UAVs, 'Figures 11 and 12 are identical.' **There is no Figure 11 or 12 in this paper, which contains only 4 figures.**"

This claim is simply false. The paper referenced here (authored by JS Chahl, MV Srinivasan, and SW Zhang and for which we included the link in our manuscript) contains 12 figures, and we are reproducing the identical figures 11 and 12 here:
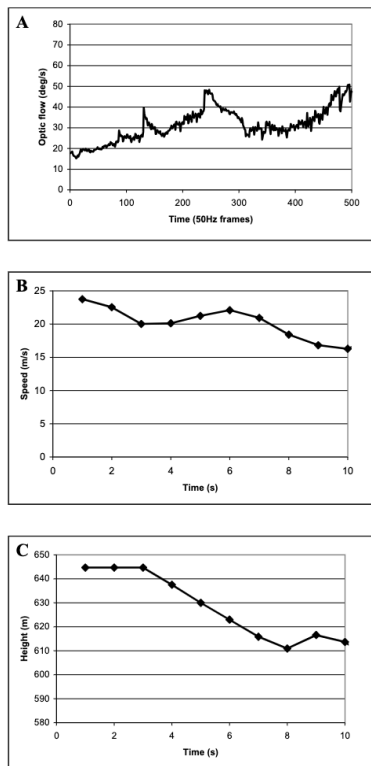

Fig. 11. The results of testing the concept of descent controlled by optic flow. Despite the decrease in altitude (C), the optic flow (A) did not increase, as the elevator was used to control air speed (B) by altering the rate at which potential energy was converted to kinetic energy. Altitude is measured here as height above sea level. The lower curve represents
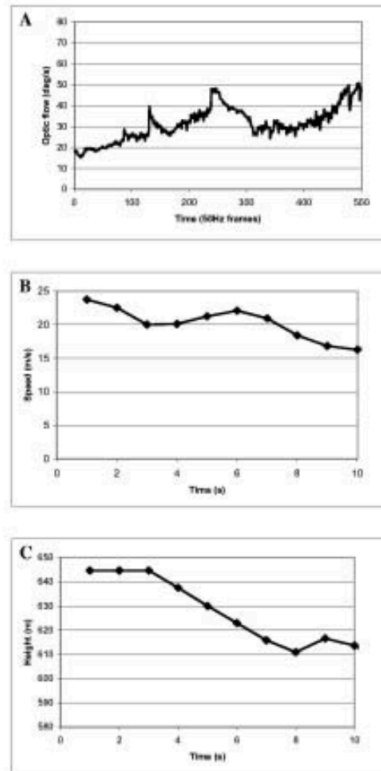
Fig. 12. The same experiment was run as in Figure 11. In this case the throttle was reduced, and the controls held fixed in their trim positions. As the altitude decreases (C) the optic flow increases (A), while speed remains constant (B), as expected, bounded by drag.

Next, Srinivasan et al. address our analysis of the repeated reporting of extremely high R^2 values. They begin with "On the question of a high linear regression coefficient reported for the graph in Fig. 2 of Srinivasan et al. (2000a) [8]: [...] **Recalculation of the regression coefficient yields a value of 0.998, which is exactly the published value of 0.998 - there is no error.**" **We never claimed that this R^2 value was calculated incorrectly.** In fact, our manuscript states directly, "the mean values reported in Srinivasan et al. (2000) are almost perfectly linear with R2 = 0.997". **Instead, we reported that the R^2 value in a different paper was reported incorrectly (Figure 1c of the 2005 paper Zhang et al.).** Anybody can reproduce our calculation and see for themselves using the code we provided with our original manuscript: https://github.com/pachterlab/LP_2024/blob/main/bee_real_analysis.ipynb.

The authors do admit that "R^2 values reported in some of the papers in question are actually R values, not R^2 values", which apparently does not constitute an error because they also claim: "to the best of our knowledge, there are no errors in the calculation of the regression coefficients."

Moreover, the authors include an Appendix B, in which R^2 values are recalculated from reported means. Again, this is a strawman argument since we never claimed (with the exception of Zhang et al. above) that the reported R^2 values do not match the reported means. Rather, **our simulations show that obtaining the reported means in combination with the reported standard deviations would have been unlikely. Srinivasan et al. either did not understand the purpose of our analysis or willfully distorts it** since they incorrectly claim: "L&P have also claimed, via simulation, that the reported R^2 value in Srinivasan et al. (2000) is not reproducable from the reported means and standard deviations in Table 1 of that paper."

3)
According to Srinivasan et al., we made an "elementary statistical error" in our simulations which is described in detail in a [report written by GW Stuart](). Stuart describes two "flaws" with our simulations.

The first "flaw" pertains to the hierarchical structure of the means including averaging over bees, dances, and waggles. Stuart basically argues that our simulation should assume that bees share the same mean, even though there is no experimental evidence for this. While we are not sure why this would be a correct assumption, we actually performed this simulation in our manuscript as noted by Stuart himself "the second simulation, relegated to Supplementary material, does not suffer from Flaw 1." From our manuscript:
"In Srinivasan et al. (2000), the data was likely averaged over individual bees, dances, and waggle phases to generate a mean for each feeder location. In performing our analysis, we assumed that bees do not share the same mean (Schürch et al., 2016). It is possible that averaging and binning, e.g. based on the camera frame rate, contributed to obtaining the high R2 values that are reported throughout these papers."

As Stuart describes, we report that - if you assume they have the same variance - "the proportion of R2 values that exceed the original value is 19%." While it is true that, as Stuart asserts, the 19% alone "is not strong evidence against the credibility of the original reported R^2" we describe in our manuscript, "taking into account both papers, we arrive at a probability of 0.19 * 0.14 = 0.0266" - and this is just two papers, this will further decrease when taking into account all six papers reporting R2 ≥ 0.99.

The second flaw pertains to the data points we used to compute our simulations. Stuart seems to refer to the means reported in the original paper as the "expected values" and the regression line computed from said means as "observed means." Not surprisingly, he finds that if you simulate the regression as if it were the reported means, you get said regression back at a very

high rate. This analysis has little to do with the analysis we performed in our original manuscript, which simulates the likelihood of the reported means given the reported standard deviations.