Matters arising

Magnetic field effects on behaviour in *Drosophila*

https://doi.org/10.1038/s41586-024-07319-x

Received: 15 November 2023

Accepted: 15 March 2024

Published online: 1 May 2024

Open access

Check for updates

Steven M. Reppert[1]✉

ARISING FROM: M. Bassetto et al. *Nature* https://doi.org/10.1038/s41586-023-06397-7 (2023)

The fruit fly *Drosophila melanogaster* is a model organism that has been used by several laboratories to study geomagnetic sensing and its molecular basis. Bassetto et al.[1] proclaim that there is no evidence for magnetic field effects on behaviour in *Drosophila*. I challenge their conclusion and defend the work in Gegear et al. 2008 (ref. 2), in which a binary-choice T-maze assay not only was used to reveal fruit fly magnetosensitivity but also provided mechanistic insights into the role of the ultraviolet-A/blue-light photoreceptor cryptochrome (Cry) in the magnetic response. Reviewing all of the published data, there is considerable evidence for magnetosensitivity in fruit flies.

Gegear et al. 2008 (ref. 2) developed a viable *Drosophila* behavioural assay for assessing magnetosensitivity at a field intensity of 500 μT. In an illuminated apparatus, flies experience a magnetic field generated by an electric coil system and exhibit their magnetosensitivity in a binary-choice T-maze[2–4]. The two-coil system is ideal for behavioural studies of magnetosensitivity, because it produces a magnetic field on one side of the T-maze, while producing no field on the opposite side. Importantly, the studies were carried out in the same laboratory where olfactory conditioning controls were routinely carried out in which flies are trained to associate odours with sugar reward. In the T-maze assay, wild-type flies showed significant naive and trained responses to the magnetic field, and the responses were light dependent. The ultraviolet-A/blue-light photoreceptor Cry[5] mediated the light-dependent magnetosensitivity. In a second study, Gegear et al. 2010 (ref. 3) showed that when a *cry* transgene is properly expressed in Cry-deficient flies, a full magnetic response with appropriate light activation is restored. All of the data discussed herein are from published resources.

Any behavioural paradigm is sensitive to the environment in which it is carried out and this is particularly the case for fly conditioning. It is arguably the most complex of these types of *Drosophila* phenotype and requires considerable skill and experience to obtain reliable results. Although the experiments of Bassetto et al.[1] might have been optimally shielded against interfering outside magnetic effects, it is evident in their Methods section that the critical 'positive conditional control' utilizing olfactory conditioning was not carried out under the same conditions as the failed magnetic conditioning studies. Instead, these 'controls' were carried out under temperature- and humidity-controlled conditions in Oxford, UK. Without 'controls' under the same location and conditions, it is impossible to determine whether the shielded location in Oldenburg, Germany, had the appropriate environment (humidity and temperature) that permits robust sugar-reinforced conditioning. The lack of an appropriate 'positive conditional control' in Oldenburg is a substantial criticism and suggests that there may be other important variables that differ between the studies in Bassetto et al.[1] and those in Gegear et al. 2008 (ref. 2).

Bassetto et al.[1] emphasize the large number of flies they tested (97,658) in the T-maze without finding a magnetic response, compared to the "small sample size" used in Gegear et al. 2008 (ref. 2). Notably, >39,500 flies were used to complete the studies in Gegear et al. 2008 (ref. 2). There were 390 groups of 100–150 flies used; the number of flies is easy to calculate from the data in the figures. This comparatively large number of flies used is in stark contrast to the small number of flies implied by Bassetto et al.[1] and in the News and Views piece by Warrant[6].

Bassetto et al.[1] next reassessed the statistical analysis in Gegear et al. 2008 (ref. 2). Their reanalysis is off base and does not support the contention that most of the original results were not statistically significant and were instead false positives.

Bassetto et al.[1] criticize the use of parametric statistical testing in the Gegear et al. 2008 paper[2]. However, analysis of *Drosophila* conditioning data is frequently carried out using parametric statistics. Indeed, Krashes and Waddell[7,8] advise using parametric statistical testing of performance index values derived from appetitive and aversive olfactory conditioning assays and recommend a sample size of 8–10 replicates per condition per genotype. Instead, Bassetto et al.[1] have selected an extremely conservative approach to reanalysis of the data in Gegear et al. 2008 (ref. 2). This choice leads to misguided conclusions on the statistical power of the original analysis.

When using an ordinal logistic fit model to assess the synthetic dataset, which is equivalent to the type of generalized linear model used by Bassetto et al.[1] (based on the group averages in Gegear et al. 2008, Fig. 1b[2]; discussed in the text and Supplementary Fig. 1a of Bassetto et al.[1]), the statistical results are very dependent on how the batches of about 100 flies (in each experiment) are encoded in the model. With 'batch' included as an independent variable, the effect of training is minimal ($P = 0.33$), whereas omission of 'batch' altogether leads to a highly significant effect of training ($P < 0.0001$). Presumably, Bassetto et al.[1] chose the former option.

Our conclusion that the approach of Bassetto et al.[1] is overly conservative is based on a much more straightforward, non-parametric approach (the Wilcoxon rank sum test, also known as the Mann–Whitney *U*-test). The data for the naive and trained groups of flies in the synthetic dataset are highly significantly different by this analysis ($P < 0.0001$).

Mimicking the approach of Bassetto et al.[1] to generate a single synthetic dataset, we generated an additional 20 synthetic datasets. When using the very conservative approach of Bassetto et al.[1] (presumably a binominal approach with 'batch' as an independent variable), 5 of

[1]Department of Neurobiology, UMass Chan Medical School, Worcester, MA, USA. ✉e-mail: steven.reppert@umassmed.edu

20 datasets demonstrated a significant effect of training, whereas 15 did not. When using three other approaches (*t*-tests, ordinal logistic binominal models without 'batch' or non-parametric rank tests), all 20 synthetic replicates demonstrated highly significant differences between the groups ($P < 0.0001$). Thus, Bassetto et al.[1] seem to have selected a statistical approach with extremely poor sensitivity for detecting differences when reanalysing the data in Gegear et al. 2008 (ref. 2). Their conclusion that the results in Gegear et al. 2008 (ref. 2) represent a 'false positive' is unfounded. Moreover, if false positives occurred in previous studies, they would be expected to occur in a variety of treatments and not in a way that consistently provides evidence for magnetosensitivity.

Bassetto et al.[1] also criticize the statistical approach used in Gegear et al. 2008 (ref. 2) by stating that it assumes independence of each fly in a batch and subsequently treats each fly as an independent biological replicate, violating the requirement for independence of the samples and leading to pseudo replication. In fact, statistical analysis was carried out on the 8–12 independent values for performance index per group (each of which was derived from an independent batch of 100–150 flies). There is no pseudo replication.

Bassetto et al.[1] were also unable to detect a magnetic effect on negative geotaxis in *Drosophila*, as reported in Fedele et al.[9]. Importantly, the magnetic response reported by Fedele et al.[9] was replicated independently by Bae et al.[10]. This replication is not mentioned by Bassetto et al.[1]. Instead, they tried but were unable to replicate the work in Fedele et al.[9]. The inability of Bassetto et al.[1] to replicate the work of not only Fedele et al.[9] but also Bae et al.[10] makes their negative results less convincing.

There are at least 15 papers over the past 50 years reporting the existence of a fly magnetic sense, and several of these suggest a Cry-based mechanism (papers listed in Bassetto et al.[1]). Most of these reports used assay systems other than the T-maze and negative geotaxis paradigms. Nevertheless, Bassetto et al.[1] dismiss all of these other reports. Their refutation of these studies without direct evidence is unsubstantiated.

Bassetto et al.[1] conclude by claiming that night-migratory songbirds (which are technically challenging for any kind of molecular genetic analyses) remain the organisms of choice for elucidating the mechanism of light-dependent magnetosensitivity. However, the authors overlooked the published work on the biologically relevant magnetic compass of the migratory monarch butterfly. Two independent reports that use distinctive behavioural assays show that individual monarchs manifest robust light-dependent inclination magnetic responses to Earth-strength magnetic fields[11,12]. Moreover, genetic studies show that the photoreceptive Cry1 protein is essential for the monarch's light-sensitive magnetic compass[12]. The recent successful use of reverse genetics in monarchs[12] indicates that the butterfly is an excellent choice for delineating the molecular mechanisms underlying light-dependent magnetosensing in the context of compass navigation.

1. Bassetto, M. et al. No evidence for magnetic field effects on the behaviour of *Drosophila*. *Nature* **620**, 595–599 (2023).
2. Gegear, R. J., Casselman, A., Waddell, S. & Reppert, S. M. Cryptochrome mediates light-dependent magnetosensitivity in *Drosophila*. *Nature* **454**, 1014–1018 (2008).
3. Gegear, R. J., Foley, L. E., Casselman, A. & Reppert, S. M. Animal cryptochromes mediate magnetoreception by an unconventional photochemical mechanism. *Nature* **463**, 804–807 (2010).
4. Foley, L. E., Gegear, R. J. & Reppert, S. M. Human cryptochrome exhibits light-dependent magnetosensitivity. *Nat. Commun.* **2**, 356 (2011).
5. VanVickle-Chavez, S. J. & Van Gelder, R. N. Action spectrum of *Drosophila* cryptochrome. *J. Biol. Chem.* **282**, 10561–10566 (2007).
6. Warrant, E. J. Replication study casts doubt on magnetic sensing in flies. *Nature* **620**, 496–497 (2023).
7. Krashes, M. J. & Waddell, S. *Drosophila* aversive olfactory conditioning. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5608 (2011).
8. Krashes, M. J. & Waddell, S. *Drosophila* appetitive olfactory conditioning. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5609 (2011).
9. Fedele, G. et al. Genetic analysis of circadian responses to low frequency electromagnetic fields in *Drosophila melanogaster*. *PLoS Genet.* **10**, e1004804 (2014).
10. Bae, J.-E. et al. Positive geotactic behaviors induced by geomagnetic field in *Drosophila*. *Molec. Brain* **9**, 55 (2016).
11. Guerra, P. A., Gegear, R. J. & Reppert, S. M. A magnetic compass aids monarch butterfly migration. *Nat. Commun.* **5**, 4164 (2014).
12. Wan, G., Hayden, A. N., liams, S. E. & Merlin, C. Cryptochrome 1 mediates light-dependent inclination magnetosensing in monarch butterflies. *Nat. Commun.* **12**, 771 (2021).

Acknowledgements I thank B. Kyriacou, D. Henshaw, S. Waddell, L. E. (Foley) North, C. Merlin and D. R. Weaver for discussions and comments on the manuscript, and V. van der Vinne for statistical analysis.

Competing interests The author declares no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Steven M. Reppert.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Matters arising

Magnetic field responses in *Drosophila*

https://doi.org/10.1038/s41586-024-07320-4

Received: 24 November 2023

Accepted: 15 March 2024

Published online: 1 May 2024

Open access

Check for updates

Charalambos P. Kyriacou[1]✉

ARISING FROM: M. Bassetto et al. *Nature* https://doi.org/10.1038/s41586-023-06397-7 (2023)

Bassetto et al.[1] reported that *Drosophila* are unable to detect magnetic fields using a conditioning[2] and negative geotaxis assay[3], and on this basis, they dismiss these and all further experimental studies published on *Drosophila* magnetic fields[4–12]. Critically, fly magnetic geotactic responses were replicated independently by Bae et al.[12], yet this important and extensive confirmatory study is not discussed. Furthermore, Bae et al. successfully demonstrated a magnetic field conditioning response[12], underlining how experienced *Drosophila* groups can successfully negotiate magnetic paradigms. I have reanalysed the data from all three geotactic experiments from Bassetto et al.[1] and, despite serious flaws in methodology, their results reveal that *Drosophila* detect magnetic fields.

In the geotaxis experiments of Fedele et al.[3], the percentage of male flies climbing 15 cm in 15 s generated the maximum separation between sham responses to blue light (BL) and those to red light (RL), which provide the critical positive controls. Forty-eight per cent of CS-LE males exposed to BL reached this criterion, compared to 12% in RL (that is, about 36% absolute, 400% relative enhancement in BL; Fig. 1a). The contention of Bassetto et al.[1] that flies should fall into either climber or non-climber categories and not reflect an underlying Gaussian distribution does not stand serious scrutiny. Bae et al.[12] carried out similar experiments, comparing geotaxis at about 0 μT magnetic field in darkness (approximately equivalent to RL for flies) and white light (500 lx, including BL). They observe a geotactic difference of about 300% between the two lighting conditions expressed as positive geotaxis (non-climbers; Fig. 1b). Experiment 1 of Bassetto et al.[1] replicates the procedure of Fedele et al.[3] in equipment I provided, but apparently using mixed groups of males and females. It is of concern that geotactic responses are barely different between the sham BL and RL critical positive controls (Fig. 1a). I recalculated that the CS-LE strain reached 26% criterion under BL with 22% under RL, whereas corresponding values for the more active CS-OX were 52% (BL) and 45% (RL) (Fig. 1a). Given these tiny absolute and relative differences between RL and BL, compared to those in previous studies[3,12], one questions how any magnetic field effect could be detected in such limited phenotypic space. Evidently, Bassetto et al.[1] did not suspect a problem with these positive controls (see Supplementary Information for the probable reason). In addition, strain CS-LE is considerably less active in BL than in Fedele et al.[3], possibly owing to inbreeding, as I originally provided a single vial of this line. Consequently, I predominantly limit my reanalyses to CS-OX, which is as active in BL as CS-LE is in Fedele et al.[3] (Fig. 1a).

In experiment 2, Bassetto et al.[1] expose groups of 10 individuals to 0 μT, at which Earth's magnetic field is neutralized, compared to 90-, 220- and 300-μT exposures with corresponding sham (ambient, about 40-μT) controls. They do not use 500-μT exposures as in Fedele et al.[3]. Inspecting the automated tracking for CS-OX revealed 1,062,956 frames logged from an expected 1,800,000 (accuracy 59%). Importantly, no positive controls were carried out involving RL versus BL for CS-OX. Nevertheless, taking their results at face value, the prediction[3,12] is that flies should climb higher at 0 μT compared to magnetic field exposure. Reanalysis of their data reveals significantly higher climbing at 0 μT than at 90-, 220- and 300-μT exposures combined (Fig. 1c). Also predicted is that 0-μT-exposed flies should climb higher than corresponding shams, but the higher-intensity exposures should reduce climbing compared to sham, generating an interaction. Figure 1e reveals that at 90-, 220- and 300-μT exposures, climbing is reduced compared to corresponding shams, as expected (but not significantly), whereas there is little difference between 0 μT compared to its sham.

For Flyvac experiment 3, Bassetto et al.[1] tracked individual flies. The accuracy of the tracking is 84.9%, considerably better than experiment 2. In CS-LE, 12.5% (26/208) of BL trials included flies that reached the climbing criterion (15 cm in 15 s in at least 1 of 5 trials), compared to 3% in RL (6/199). The mean percentage of flies across all trials reaching criteria was 4% for BL and 1% for RL, so this criterion cannot be used to investigate magnetic field effects (Fig. 1a). Nevertheless, I detected significant differences between the lighting conditions using Fisher exact ($P = 0.004$) and Mann–Whitney ($P = 0.001$) tests, reflecting absolute BL-to-RL enhancement of 9.5%, relative 415%. Consequently, I recalculated the mean height climbed at 15 s for CS-LE under sham in RL and BL, which was 6.17 cm to 8.75 cm (142% BL enhancement), considerably better than experiment 1. Yet again, positive RL and BL controls were not carried out for CS-OX, so I assumed that CS-OX discriminates BL and RL as well as CS-LE does. I therefore took the average height climbed for CS-OX individuals at 15 s and reanalysed the data. The prediction is that 0-μT-exposed flies should climb higher than those of the other exposures combined. The prediction is partially fulfilled, but unlike experiment 2, the difference is not significant (Fig. 1d). Flies exposed to 0 μT should also climb higher in BL than sham (about 40 μT), but at 300 μT, sham flies should climb higher than exposed flies. Two-way analysis of variance (ANOVA) reveals a significant interaction generated by the flies at 0 μT climbing higher than sham, with a strong reciprocal significant response at 300 μT (Fig. 1f), so the prediction is fulfilled. A similar result is obtained when I examined the percentage of CS-OX flies reaching criterion (15 cm in 15 s), noting how much higher CS-OX climb than CS-LE in BL (compare experiment 3 in Fig. 1a with Fig. 1g). One wonders what the result would have been had Bassetto et al.[1] used an exposure of 500 μT (as in Fedele et al.[3]), as the magnetic field effect in this particular single-fly paradigm seems to gain momentum with increasing intensity. The over-elaborate and highly conservative ANOVA of Bassetto et al.[1] (see Methods of ref. 1) produced non-significant results, after which the authors did not seem to interrogate their data further. Had they inspected carefully the relevant part of their own

[1]Neurogenetics Group, Department of Genetics & Genome Biology, University of Leicester, Leicester, UK. ✉e-mail: cpk@leicester.ac.uk

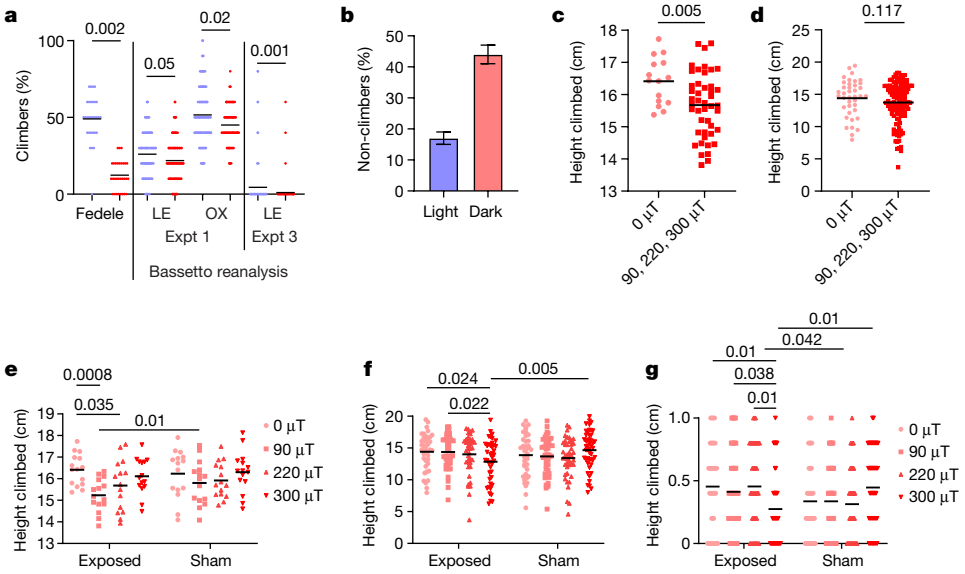

Fig. 1 | Results of reanalysis of geotaxis data in Bassetto et al.[1]. Raw data are shown for all experiments; horizontal black lines represent means. **a**, Reanalysis of Bassetto et al.[1] raw data for positive control conditions. The plot shows a comparison of raw data for climbing response of CS-LE flies under sham to RL and BL (red and blue, respectively) conditions from Fedele et al.[3] with raw data for CS-LE and CS-OX (experiment (Expt) 1) reanalysed from Bassetto et al.[1]. The *y* axis shows the percentage of flies that reached 15 cm in 15 s. Mean climbing scores averaged over 10 trials for each tube for the experiment of Fedele et al.[3] (one-tailed $t_4 = 5.82$, $P = 0.002$, based on 3 replications each for RL and BL; total observations, $n = 60$). Experiment 1 of Bassetto et al.[1] has 300 observations under sham, divided equally between BL and RL, in which 60 tubes (each with 10 flies) are tested 5 times. The raw data and mean responses are shown for CS-LE and CS-OX. For ANOVA, the proportion of flies reaching criterion is calculated for each tube (strain: $F_{1,117} = 17.07$, $P \ll 0.0001$; light: $F_{1,117} = 9.82$, $P = 0.002$; interaction: $F = {}_{1,117}\,0.09$, not significant (NS); based on $n = 121$ average climbing scores from 5 trials (total flies, $n = 601$)). False discovery post hoc values are shown (see Supplementary Information). In experiment 3 of Bassetto et al.[1], only 26/208 (12.5%) and 6/199 (3%) CS-LE trials produced flies that reached criterion in BL and RL, respectively, so 182 and 193 trials, respectively, had a score of 0. The mean percentage of five trials in which individual flies reached criterion and Mann–Whitney *U*-test result comparing BL to RL are shown. It is clear from the raw data that there is barely any overlap between RL and BL climbing scores in the positive controls of Fedele et al.[3], whereas the overlap is considerable in the raw data for experiments of Bassetto et al.[1]. **b**, Results of the experiment of Bae et al.[12] comparing climbing in darkness and white light at 0 µT. Redrawn from Bae et al.[12]; raw data not available. Data are mean ± s.e.m. **c**, Reanalysis of climbing of 0-µT-exposed CS-OX flies in groups of 10 individuals compared to higher 90-, 220- and 300-µT exposures from gravity experiment 2 of Bassetto et al.[1] (one-tailed $t_{58} = 2.64$, $P = 0.005$, $n = 60$). **d**, Same analysis and comparison for Flyvac experiment 3 of Bassetto et al.[1], in which individual CS-OX flies are tracked (one-tailed $t_{160} = 1.19$, $P = 0.117$, $n = 162$). **e**, Reanalysis of gravity experiment 2 with CS-OX from Bassetto et al.[1]. Mean height (horizontal bar) climbed per tube in 15 s. ANOVA, exposure versus sham: $F_{1,112} = 1.42$, NS; exposure intensity: $F_{3,112} = 4.67$, $P = 0.004$; interaction: $F_{3,112} = 0.8$, NS; $n = 120$. False discovery post hoc *P* values shown. **f**, Reanalysis of Flyvac experiment 3 with CS-OX from Bassetto et al.[1]. Mean height climbed per tube in 15 s. Exposure versus sham: $F_{1,333} \approx 0$, NS; exposure intensity: $F_{3,333} = 0.38$, NS; interaction: $F_{3,333} = 3.44$, $P = 0.017$; $n = 341$; false discovery post hoc *P* values shown. **g**, Proportion of flies that reached criterion of 15 cm in 15 s from Flyvac experiment 3. The horizontal bar depicts the mean. ANOVA, exposure versus sham: $F_{1,348} = 1.47$, NS; exposure intensity: $F_{3,348} = 0.19$, NS; interaction: $F_{3,348} = 4.40$, $P = 0.0036$, $b = 356$; false discovery post hoc *P* values shown. **b**, Redrawn from ref. 12, Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

figures (see my Supplementary Fig. 1), they might have thought twice about their conclusions.

I have shown that the positive controls for experiment 1 worked poorly, if at all, and that in experiment 2, comparing 0-µT exposures to the higher exposures gave the expected result, despite poor tracking accuracy and no positive controls. In the more robust final experiment, despite no positive controls, the interaction expected, in which flies climb higher under 0 µT and lower under higher exposures compared to sham, also gave the predicted result. Instead of engaging in some relatively simple troubleshooting for each paradigm, increasing BL intensity in experiment 1 (and perhaps experiments 2 and 3), and tuning up the tracking software in experiment 2, Bassetto et al.[1] preferred the option of simply racking up large (108,609) numbers. It is extraordinary that no positive RL or BL controls were carried out for CS-OX, because it has long been known that fly strains differ in their responses to RL[13].

Finally, one wonders why Bassetto et al.[1] dismissed all fly magnetic field experiments[2–12] from eight independent groups using different paradigms. Bassetto et al.[1] state that because flies do not use a navigational compass, they have no use of a magnetic sense. They ignore the demonstration of Bae at al.[12] that flies use the Earth's magnetic field to fly low. *Drosophila melanogaster* feed and oviposit on decaying fruits that lie mainly at ground level, so a magnetic sense would be adaptive for foraging. In turn, this suggests that magnetoreception is primary, and the functions it serves, foraging or navigation, lie downstream. Furthermore, magnetic field effects can be mediated in flies by the 52-residue cryptochrome (Cry) carboxyl terminus alone without the canonical FAD-binding site and the 3–4 Trp residues required to generate radical pairs in Cry, results obtained using adult circadian behaviour (under impeccably controlled conditions) and single-larval-motoneuron physiological assays[8,10,11]. Mouritsen, Hore and collaborators favour a model in which full-length avian CRY4 with FAD binding and Trp tetrads is required for detecting magnetic fields, based on in vitro spectroscopy experiments on CRY4 peptides circumstantially allied to behavioural evidence from bird navigation studies[14]. Clearly the two competing hypotheses, Cry C terminus versus full-length Cry, although not mutually exclusive, are at odds. The critically flawed attempt of Bassetto et al.[1] to cast doubt on all fly magnetic field work, together with their statement that (genetically and molecularly inscrutable) night-migratory songbirds are the best organism for understanding the underlying mechanism of light-dependent magnetoreception (ignoring the molecularly tractable navigating monarch butterfly[15]), should be seen clearly in this context.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All of the raw data for Bassetto et al.[1] on which the analyses are based can be found at Open Science Framework (https://doi.org/10.17605/OSF.IO/HZ98Q).

1. Bassetto, M. et al. No evidence for magnetic field effects on the behaviour of *Drosophila*. *Nature* **620**, 595–599 (2023).
2. Gegear, R. J., Casselman, A., Waddell, S. & Reppert, S. M. Cryptochrome mediates light-dependent magnetosensitivity in *Drosophila*. *Nature* **454**, 1014–1018 (2008).
3. Fedele, G., Green, E. W., Rosato, E. & Kyriacou, C. P. An electromagnetic field disrupts negative geotaxis in *Drosophila* via a CRY-dependent pathway. *Nat. Commun.* **5**, 4391 (2014).
4. Gegear, R. J., Foley, L. E., Casselman, A. & Reppert, S. M. Animal cryptochromes mediate magnetoreception by an unconventional photochemical mechanism. *Nature* **463**, 804–807 (2010).
5. Wehner, R. & Labhart, T. Perception of geomagnetic field in fly *Drosophila melanogaster*. *Experientia* **26**, 967–968 (1970).
6. Dommer, D. H., Gazzolo, P. J., Painter, M. S. & Phillips, J. B. Magnetic compass orientation by larval *Drosophila melanogaster*. *J. Insect Physiol.* **54**, 719–726 (2008).
7. Yoshii, T., Ahmad, M. & Helfrich-Forster, C. Cryptochrome mediates light-dependent magnetosensitivity of *Drosophila*'s circadian clock. *PLoS Biol.* **7**, 813–819 (2009).
8. Fedele, G. et al. Genetic analysis of circadian responses to low frequency electromagnetic fields in *Drosophila melanogaster*. *PLoS Genet.* **10**, e1004804 (2014b).
9. Wu, C.-L. et al. Magnetoreception regulates male courtship activity in *Drosophila*. *PLoS ONE* **11**, e0155942 (2016).
10. Giachello, C. N. G., Scrutton, N. S., Jones, A. R. & Baines, R. A. Magnetic fields modulate blue-light-dependent regulation of neuronal firing by cryptochrome. *J. Neurosci.* **36**, 10742–10749 (2016).
11. Bradlaugh, A. A. et al. Essential elements of radical pair magnetosensitivity in *Drosophila*. *Nature* **615**, 111–116 (2023).
12. Bae, J.-E. et al. Positive geotactic behaviors induced by geomagnetic field in *Drosophila*. *Molec. Brain* **9**, 55 (2016).
13. Kyriacou, C. P. & Burnet, B. Genetic analysis of phototaxis near the upper limit of the visual spectrum of *Drosophila melanogaster*. *Behav. Genet.* **9**, 123–128 (1979).
14. Xu, J. et al. Magnetic sensitivity of cryptochrome 4 from a migratory songbird. *Nature* **594**, 535–540 (2021).
15. Wan G., Hayden A. N., Liams S. E. & Merlin C. Cryptochrome 1 mediates light-dependent inclination magnetosensing in monarch butterflies. *Nat. Commun.* **12**, 771 (2021).

Acknowledgements I thank M. Bassetto and M. Winklhoefer for discussions about their data, and A. Jones for comments on the manuscript. I acknowledge the support of BBSRC grant BB/V006304/1.

Competing interests The author declares no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07320-4.

Correspondence and requests for materials should be addressed to Charalambos P. Kyriacou.

nature portfolio

Corresponding author(s): C. P. Kyriacou

Last updated by author(s): Feb 29, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	data collected by Bassetto et al https://doi.org/10.17605/OSF.IO/HZ98Q
Data analysis	GraphPad Prism

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data availability. All the Bassetto et al raw data on which the analyses are based can be found at https://doi.org/10.17605/OSF.IO/HZ98Q

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Bassetto et al are unclear about the sex of flies used. Only in the gravity geotactic experiment do they specifically state that male flies were used. In the Bassetto PhD thesis he states that for the third experiment (FlyVac) 'the same gender was used' as in the gravity experiment so I assume they used males too. In experiment 1, the direct replication of Fedele et al, they do not state anything about gender, so I presume male and female flies were used. This is unfrrtunate as male and female flies will interact in the climbing assay,thereby adding another source of uncontrolled variation.
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All the reanlayses are based on the large sample sizes of Bassetto et al. The degrees of freeedom and sample sizes used in the statistical tests are provided
Data exclusions	For Flyvac experiment 3 Bassetto took 5 trials for each individual male. However 8 males generated data for only 1 trial so an average climbing score across trials could not be computed. Nevertheless, the data was analyses both with and without these 8 datapoints, but the statistical results were the same (see Supplementary Methods)
Replication	This study represents a re-analysis of data by Bassetto et al who could not replicate the results of Fedele et al (2014). The re-analysis reveals that in fact the Bassetto et al data support the conclusions of Fedele et al, so the initial non-replication by Bassetto et al appears to be flawed.
Randomization	Basssetto et al randomised the order of sham and exposure
Blinding	Bassetto et al apparently performed their experiments blind

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Drosophila melanogaster
Wild animals	n/a
Reporting on sex	Only one geotactic experiment out of three reports the sex used. Another probably used males (FlyVac experiment 3), but no mention is made for the sex of that animals in experiment one, the direct replication of Fedele et al
Field-collected samples	n/a
Ethics oversight	none required for Drosophila

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Bassetto et al. reply

https://doi.org/10.1038/s41586-024-07321-3

Published online: 1 May 2024

Open access

Check for updates

Marco Bassetto[1,2], Thomas Reichl[2], Dmitry Kobylkov[2,3], Daniel R. Kattnig[4,5], Michael Winklhofer[6,7]✉, P. J. Hore[1]✉ & Henrik Mouritsen[2,7]✉

REPLYING TO: S. M. Reppert *Nature* https://doi.org/10.1038/s41586-024-07319-x (2024)

REPLYING TO: C. P. Kyriacou *Nature* https://doi.org/10.1038/s41586-024-07320-4 (2024)

We welcome the opportunity to respond to two Comments[1,2] on our study[3] of the behaviour of fruit flies in magnetic fields. In Bassetto et al.[3], we attempted to implement the assays of Gegear and colleagues[4–6] and Fedele et al.[7], hoping that they would allow us to use *Drosophila* as a model organism for determining the biophysical mechanisms, genetic basis and neuronal pathways by which animals respond to magnetic stimuli. This proved to be a fruitless endeavour.

Our conclusion[3] that the magnetic field effects reported by Gegear et al.[4] were most likely false positives was based on the incorrect choice of statistical tests by these authors. We have discussed these matters extensively in the Supplementary Information of our paper[3] and in ref. 8. Here we provide only a short summary.

The Student's *t*-test (and similarly analysis of variance) proposed by Krashes and Waddell[9] and used by Gegear et al.[4] to analyse group T-maze data is strongly affected by pseudo-replication and is fundamentally wrong for analysing preference indices[3,8]. This is reflected in the exaggeratedly significant results ($P < 0.0001$) claimed in ref. 4 for small proportion contrasts (45% naive versus 55% trained). We therefore chose a correct statistical framework for proportions and avoided pseudo-replication by taking each batch of flies as the independent statistical replication unit (biological replicate). This correction, albeit conservative, nonetheless yielded significant results in our positive control experiment (odour-conditioned flies). Reppert[2] includes additional statistical tests in his Comment (ordinal logistic fit model and Wilcoxon rank sum test) and presents results derived from synthetic data. Ignoring the pseudo-replicative nature of the data, these tests suffer from the same problem as the *t*-test (see above). Statistical tests are based on frameworks of assumptions that are appropriate for specific problems and data structures and cannot be applied out of context. We cannot comment further on these analyses because neither the original data[4] nor the new synthetic data[2] have been made available.

Reppert[2] also writes that the fact that we did not attempt to replicate the negative geotaxis experiments of Bae et al.[10], together with our failure to reproduce the findings of Fedele et al.[7], makes our study[3] an outlier. Once again, this opinion ignores the fact that these studies used statistical approaches that are not appropriate for proportions and therefore led to highly exaggerated *P* values for small to moderate proportion contrasts. Both Fedele et al.[7] and Kyriacou[1] (Fig. 1b[1]) attach inappropriate and deceptive s.e.m.-based error bars to the proportion of non-climbers and thus greatly underestimate the uncertainty in the proportions due to pseudo-replication. Applying the *t*-test to proportions implicitly assumes that each fly in a batch is an independent biological replicate in the extremely strict sense that the decision of each individual was interrogated independently of the other flies (that is, as if each fly were tested individually)[8]. It is clear that the use of an intrinsically pseudo-replicative rapid group assay makes it impossible to know how many flies in a batch made a decision independently of the others (see also Mora et al.[11]). This is why we consider a batch of flies, rather than an individual fly, to be an independent biological replicate. In conclusion, a larger number of studies using invalid statistics does not make them more convincing.

Kyriacou[1] reanalyses selected parts of our negative geotaxis data, reaching different conclusions. As in Fedele et al.[7], the first part of his reanalysis is based on an arbitrary criterion as to the definition of climbers, without providing his own data to demonstrate that a group of flies is clearly separable into climbers versus non-climbers and that this categorization can be consistently observed in repeat trials on the same group. As demonstrated in Bassetto et al.[3], the distributions of heights climbed do not show the bimodal structure that would be required for the approach of Fedele et al.[7] to be valid[8]. The second part of Kyriacou's reanalysis deals with the mean height climbed in a given time. For example, his Fig. 1e[1] hints at a magnetic field effect at 0 μT relative to 90-μT, 220-μT or 300-μT exposures. However, Kyriacou's analysis[1] showed no effect for true magnetic field exposure versus sham exposure, the latter being the negative control matched to each magnetic field exposure, with the same currents flowing through the coils as in the magnetic field exposure, but in antiparallel directions, thereby cancelling the coil field and leaving the ambient field. Similarly, the interaction between exposure and exposure intensity was not significant either. Overall, this tallies with our analysis (Supplementary Table 10 in Bassetto et al.[3]), reporting no effects other than occasional random fluctuations, which are expected when comparing many conditions. Similarly, when Kyriacou[1] chooses a single time point (again the arbitrary 15 s) in a single dataset (Fig. 1f,g[1] for CS-OX in the FlyVac setup at 300 μT), one may see a small effect, but not when taking into account the full time-dependence of the climbing behaviour (Supplementary Table 11 in Bassetto et al.[3]). Thus, Kyriacou's Comment[1] focuses on an outlier confined to an arbitrary time point in a single exposure condition. By contrast, our statistical analyses took into consideration all of the experimental conditions together with the complete climbing performance. It is clear that any experiment has outliers, the absence of which would be suspicious. Critically, an outlier cannot be taken as an effect once the exploratory data analysis has already been carried out (a practice known as HARK-ing—hypothesizing after the results are known), but can serve only as the basis of a test hypothesis to be confirmed or rejected in a repeat experiment.

Kyriacou[1] remarks on the accuracy of the video tracking of fly movements in our negative geotaxis experiments[3]. The number of frames

[1]Department of Chemistry, Physical & Theoretical Chemistry Laboratory, University of Oxford, Oxford, UK. [2]AG Neurosensory Sciences/Animal Navigation, Institut für Biologie und Umweltwissenschaften, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany. [3]Center for Mind/Brain Science, University of Trento, Rovereto, Italy. [4]Living Systems Institute, University of Exeter, Exeter, UK. [5]Department of Physics, University of Exeter, Exeter, UK. [6]AG Sensory Biology of Animals, Institut für Biologie und Umweltwissenschaften, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany. [7]Research Center for Neurosensory Sciences, University of Oldenburg, Oldenburg, Germany. ✉e-mail: michael.winklhofer@uni-oldenburg.de; peter.hore@chem.ox.ac.uk; henrik.mouritsen@uni-oldenburg.de

logged is simply a result of flies being out of bounds. Frames in which flies were hidden by the stoppers at the top or the supports at the base of the tubes were not included in the analysis. We also did not log flies that had arrived at the top of the tubes and had started to descend. This does not imply that flies were not tracked while they were climbing. Moreover, it is not necessary to track a fly in every frame to determine its climbing rate. By contrast, Fedele et al.[7] simply reported the proportion of flies that climbed to an arbitrarily chosen height within an arbitrarily chosen time period with no further data or photographic documentation.

Kyriacou[1] suggests that the absence of a magnetic response in our direct replication of Fedele et al.[7] (using the original equipment he had kindly loaned us) was due to our low blue-light intensity's resulting in small differences in geotactic responses between red- and blue-light conditions. As mentioned in Bassetto et al.[3], we used the same intensity (0.25 μW cm^{-2}) as Fedele et al.[7]. The *P* value for the effect of the wavelength of the light on climbing performance was very highly significant for the CS-OX strain[3] and just significant for the CS-LE flies[3]. We agree that the CS-LE strain received from Kyriacou may not have been ideal, but the CS-OX strain clearly passed the control.

Kyriacou[1] wonders why we used magnetic fields weaker than those of Fedele et al.[7] (500 μT) in some of our experiments. In our exact replication of Fedele et al.[7], with Kyriacou's original apparatus, we used 500 μT. Having failed to find a magnetic response under those conditions, we used two improved experimental designs (gravity and FlyVac assays)[3] with much more homogeneous magnetic fields of up to 300 μT. The radical pair mechanism[12] provides no theoretical reason to expect a large difference in the responses to such similar field strengths.

Reppert[2] berates us for conducting the positive conditioning (olfactory) controls and the magnetic exposure experiments in different locations, claiming that sugar-reinforced conditioning is a complex behavioural paradigm, sensitive to temperature and humidity. There is no evidence in Gegear et al.[4] to support such a contention. We chose to carry out the olfactory controls[3] in Scott Waddell's laboratory in Oxford specifically to take advantage of his facilities and expertise working with odour stimuli. For similar reasons, we carried out all of the magnetic stimulus tests[3] in Oldenburg, where the experimental facilities for controlling magnetic fields are second to none[13,14].

Reppert[2] is concerned about our comments on the sample sizes used by Gegear et al.[4] made in the context of the inappropriate statistical methods used by these authors (see above). The numbers are as follows. For the wild-type Canton-S flies, which showed the strongest magnetic responses reported in their paper, Gegear et al.[4] (Fig. 1b[4]) studied 22 groups of 100–150 flies (12 trained, 10 naive), whereas Bassetto et al.[3] (Fig. 1a,b[3]) used 300 groups of about 100 flies (50 trained and 100 naive for each of the OX and LE wild-type strains). In this key experiment, using an order of magnitude more samples, we failed to find a magnetic field effect[3] even when we used the inappropriate statistical analysis used by Gegear et al.[4].

Reppert[2] regrets that we did not consider the monarch butterfly as a model organism for studying the mechanism of light-dependent magnetoreception. Given the reports that monarchs should be able to orient in the Earth's magnetic field (reviewed in ref. 15), these genetically tractable insects could be a potential alternative to *Drosophila*. However, in two separate studies[16,17], we have found no evidence that monarchs have such an ability: 140 migratory monarch butterflies tested with access to only natural geomagnetic field cues showed random orientation, whereas monarchs tested with celestial cues showed a clearly directed southwest orientation[16]. Furthermore, monarchs first flown in the normal magnetic field did not react to a horizontal 120° turn of the field even when they were kept flying in the rotated field for up to 2 h (ref. 17).

Independent replication of experimental data is the 'gold standard' in science. Meticulously carried out replication studies that fail to confirm earlier results are just as important for the integrity of knowledge as those that do. We suspect that many negative replication attempts are never published. Authors can be reluctant to write them up (and some editors to publish them), resulting in an unbalanced body of literature. We encourage anyone who has tried and failed to observe *Drosophila* magnetoreception to submit their findings to reputable journals.

1. Kyriacou, C. P. Magnetic field responses in *Drosophila*. *Nature* https://doi.org/10.1038/s41586-024-07320-4 (2024).
2. Reppert, S. M. Magnetic field effects on behaviour in *Drosophila*. *Nature* https://doi.org/10.1038/s41586-024-07319-x (2024).
3. Bassetto, M. et al. No evidence for magnetic field effects on the behaviour of *Drosophila*. *Nature* **620**, 595–599 (2023).
4. Gegear, R. J., Casselman, A., Waddell, S. & Reppert, S. M. Cryptochrome mediates light-dependent magnetosensitivity in *Drosophila*. *Nature* **454**, 1014–1018 (2008).
5. Gegear, R. J., Foley, L. E., Casselman, A. & Reppert, S. M. Animal cryptochromes mediate magnetoreception by an unconventional photochemical mechanism. *Nature* **463**, 804–807 (2010).
6. Foley, L. E., Gegear, R. J. & Reppert, S. M. Human cryptochrome exhibits light-dependent magnetosensitivity. *Nat. Commun.* **2**, 356 (2011).
7. Fedele, G., Green, E. W., Rosato, E. & Kyriacou, C. P. An electromagnetic field disrupts negative geotaxis in *Drosophila* via a CRY-dependent pathway. *Nat. Commun.* **5**, 4391 (2014).
8. Winklhofer, M. Pseudoreplication and inappropriate statistical tests in the analysis of preference index data from the T-maze group assay for *Drosophila* behaviour. Preprint at *bioRxiv* https://doi.org/10.1101/2023.12.15.571933 (2023).
9. Krashes, M. J. & Waddell, S. *Drosophila* appetitive olfactory conditioning. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5609 (2011).
10. Bae, J.-E. et al. Positive geotactic behaviors induced by geomagnetic field in *Drosophila*. *Molec. Brain* **9**, 55 (2016).
11. Mora, C. V., Davison, M. & Walker, M. M. Conditioning as a technique for studying the sensory systems involved in animal orientation, homing and navigation - a review. *J. Navig.* **62**, 571–585 (2009).
12. Xu, J. et al. Magnetic sensitivity of cryptochrome 4 from a migratory songbird. *Nature* **594**, 535–540 (2021).
13. Schwarze, S. et al. Weak broadband electromagnetic fields are more disruptive to magnetic compass orientation in a night-migratory songbird (*Erithacus rubecula*) than strong narrow-band fields. *Front. Behav. Neurosci.* **10**, 55 (2016).
14. Engels, S. et al. Anthropogenic electromagnetic noise disrupts magnetic compass orientation in a migratory bird. *Nature* **509**, 353–356 (2014).
15. Merlin, C. Insect magnetoreception: a Cry for mechanistic insights. *J. Comp. Physiol. A* **209**, 785–792 (2023).
16. Stalleicken, J. et al. Do monarch butterflies use polarized skylight for migratory orientation? *J. Exp. Biol.* **208**, 2399–2408 (2005).
17. Mouritsen, H. & Frost, B. J. Virtual migration in tethered flying monarch butterflies reveals their orientation mechanisms. *Proc. Natl Acad. Sci. USA* **99**, 10162–10166 (2002).

Acknowledgements This work was supported by the Air Force Office of Scientific Research (Air Force Materiel Command, USAF award number FA9550-14-1-0095, to P.J.H. and H.M.); the European Research Council (under the European Union's Horizon 2020 research and innovation programme, grant agreement number 810002, Synergy Grant 'QuantumBirds', to P.J.H. and H.M.); the Office of Naval Research Global (award number N62909-19-1-2045, to P.J.H.); Deutsche Forschungsgemeinschaft (SFB 1372, 'Magnetoreception and navigation in vertebrates', project number 395940726, to H.M., M.W. and P.J.H., and GRK 1885, 'Molecular basis of sensory biology', to H.M. and M.W.).

Author contributions H.M., M.W. and P.J.H. wrote the manuscript with input from M.B., D.K., D.R.K. and T.R.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Michael Winklhofer, P. J. Hore or Henrik Mouritsen.