




An undergraduate classroom experiment illustrates an effect of observer bias on data collection in animal behaviour

Thomas A. Keaney ^{a, b, c, *} , Thérésa M. Jones ^a , Raoul A. Mulder ^{a, d} 

^a School of BioSciences, University of Melbourne, Victoria, Australia

^b Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg Universität Mainz, Mainz, Germany

^c Institute of Quantitative and Computational Biosciences (IQCB), Johannes Gutenberg Universität Mainz, Mainz, Germany

^d Centre for the Study of Higher Education, University of Melbourne, Victoria, Australia

ARTICLE INFO

Article history:

Received 9 October 2023

Initial acceptance 1 December 2023

Final acceptance 26 February 2024

Available online 25 April 2024

MS. number: 23-00534R

Keywords:

blind data recording
confirmation bias
foraging
reproducibility
teaching

Behavioural ecologists frequently collect data that involve the potential for subjective judgement, so it is important that researchers in the field develop awareness of potential issues around bias in data collection. We report the results of an undergraduate classroom experiment in which we estimated the potential for students' a priori expectations to bias their estimates of behaviour. Prior to conducting a set of behavioural observations on a video-recorded flock of foraging pigeons, we randomly primed half of the students to expect the pigeons to be hungry, while the other half were primed to expect the pigeons to be satiated. Students were blind to the treatment and subsequently estimated two variables expected to differ in their potential for subjectivity: the proportion of birds in the flock that were feeding (potentially subjective), and the feeding (peck) rates of two focal individuals (potentially objective). Surprisingly, we found no evidence that observer bias affected the estimate of the percentage of birds foraging. By contrast, we found a large effect of observer bias on feeding rate estimates, with students who expected a hungry state recording inflated feeding rate estimates relative to those that expected a satiated state. We furthermore found that students' expectations of foraging state did not always match their allocated primers: bias was predicted by 'expected state' but not by 'allocated state'. Our experiment illustrates that bias associated with expectation can influence results. Furthermore, a variable we initially expected to be relatively objective proved to have a strong subjective element, inflating the effect of confirmation bias on estimation. We recommend blind data recording even when response variables are thought to be objective, as well as explicit teaching of university students about the potential for bias in data collection, and our experiment suggests a potentially useful way in which this can be undertaken.

© 2024 The Authors. Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Observational bias is a pervasive problem affecting estimation in empirical research (Balph & Balph, 1983; Kaptchuk, 2003; MacCoun, 1998). One form is confirmation bias, where the pre-conceived expectations of a researcher lead to biased measurements of response variables, skewed in the direction of the researcher's expectations (Nickerson, 1998). The degree to which confirmation bias affects estimation is hypothesized to be high when a priori expectations are strong, the measured variable in question is highly subjective and/or the variable is difficult to measure (Balph & Balph, 1983; Hirst et al., 2014). Each of these conditions is commonly present in studies of the behaviour of individuals, such as those foundational to the fields of psychology, medicine and behavioural ecology. For example, behavioural

ecologists interested in the benefits of group living may have to discriminate between predator vigilance and all other forms of behaviour, oftentimes from a distance, and/or from an obstructed viewpoint (e.g. in ostriches or impala; Bertram, 1980; Van Deventer & Shrader, 2021). Empiricism in these fields is therefore particularly likely to be affected by observer biases.

Indeed, observer bias has been found to affect data quality in behavioural ecology research (Traniello & Bakker, 2015). One set of studies has taken advantage of the minimizing effect that 'blind' data recording has on observer bias. Blinding acknowledges that experimenter biases exist, but, by concealing the treatment identities of experimental units, removes (or if imperfect, reduces) the opportunity for these biases to affect data collection/collation (Schulz & Grimes, 2002). Studies following this approach typically meta-analyse the difference in the expression of a common trait or effect size following experimental manipulation between

* Corresponding author.

E-mail address: tkeaney@uni-mainz.de (T. A. Keaney).

experiments using blinded and nonblinded designs. If observer bias affects estimation, then effect sizes in nonblind studies should be larger and more often positive in the direction of the hypothesis under investigation. For example, [van Wilgenburg and Elgar \(2013\)](#) focused specifically on studies assessing the presence of nest mate recognition in ants, and found that the occurrence of aggressive encounters (indicating a lack of nestmate recognition) was roughly doubled in studies that were conducted blind. [Holman et al. \(2015\)](#) demonstrated that among a set of studies testing hypotheses in evolutionary biology, nonblinded experiments reported effect sizes 27% larger in the hypothesized direction than blind experiments. Several other large meta-analyses have reported similar results in community ecology ([Kozlov et al., 2014](#); [Zvereva & Kozlov, 2019](#)) and epidemiology ([Bello et al., 2014](#); [Hróbjartsson et al., 2012](#); [Schulz et al., 1995](#); [Wood et al., 2008](#)), although others found no evidence for observer bias ([Crossley et al., 2008](#); [Hirst et al., 2014](#)). However, comparisons between blind and nonblind studies provide only indirect, associative evidence for observer bias.

Testing for a causal effect of bias on estimation can be done with manipulative experiments, where observers are treated with some stimulus that primes them to expect a certain outcome prior to data collection. Most experiments of this nature pertaining to animal behaviour research have focused on the measurement of overtly subjective response variables ([Marsh & Hanlon, 2004](#); those most likely to be affected by bias; [Rosenthal & Fode, 1963](#); [Tuytens et al., 2014](#)). Observers primed to expect rats to perform poorly in Y maze trials estimated reduced performance relative to observers primed to expect high performance from rats sourced from the same population ([Rosenthal & Fode, 1963](#)). A more recent experiment found that observers estimated that livestock performed more stress-related behaviour if they believed the animals had been raised in stressful conditions, relative to observers that observed the same animals, but had no such expectation ([Tuytens et al., 2014](#)). Observer bias thus appears to have a strong effect on data collection when data are on the subjective side of the subjectivity – objectivity spectrum. While there is some awareness of observer bias in behavioural ecology research (e.g. [Holman et al., 2015](#); [Traniello & Bakker, 2015](#); [van Wilgenburg & Elgar, 2013](#)), specific actions to minimize its effect, such as blind data recording, are rare (relatively recent reviews suggest that less than 15% of empirical papers published in life science journals report blind data recording; [Burghardt et al., 2012](#); [Holman et al., 2015](#); [Kardish et al., 2015](#)).

One solution to this problem is to educate future scientists about the potential for bias to affect the data they gather and suggest ways in which its effect can be reduced. In support of this solution, efforts to raise awareness of bias among established researchers, through conference workshops or voluntary training modules, have proven qualitatively effective ([Hannah & Carpenter-Song, 2013](#); [Tuytens et al., 2016](#)). Better still, if observational bias around data collection is highlighted within undergraduate curricula (as recommended for medical education; [Mincey, 2021](#)), new researchers should then be able to design experiments that include measures to mitigate the effect of observational bias, immediately upon entering the field.

In this study we aimed to estimate how prevalent observer bias was across varying points of the subjectivity – objectivity spectrum in a cohort of undergraduate science students by means of an experiment involving a simple hidden ‘primer’ treatment. Students conducted behavioural observations with varying potential for subjectivity on a video recording of foraging pigeons, *Columba livia*, after being primed to expect differing hunger states in the birds. We predicted that those observers that were primed with a ‘hungry forager’ scenario would produce inflated estimates of foraging behaviour, relative to the estimates of the ‘satiated forager’ primed

group. We also tested the expectation that behavioural metrics with greater potential for subjectivity would show stronger treatment effects.

METHODS

Experimental Design

We tested whether the a priori expectations of students biased their observations by conducting a classroom experiment annually between 2021 and 2023. Each year between 58 and 62 undergraduate student observers (in their third year at university) participated in the experiment during a workshop on unconscious bias, which they undertook as part of a subject we taught on animal behaviour. Students participating in the subject were required to complete various prerequisite subjects in zoology and ecology, and thus had basic training in observational data collection. To experimentally influence the preconceptions of observers, we primed one treatment group of observers with a ‘hungry forager’ scenario and a second group of observers with a ‘satiated forager’ scenario. Students were primed by randomly assigning them to one of two ‘primer’ treatments which described a scenario in relation to a flock of foraging pigeons. Each of the primer texts was identical in length (60 words), but one primer suggested the pigeons were near starvation, while the other suggested the pigeons were overfed (full email contents for each scenario are available in the [Appendix](#)). Students were sent individual emails 24 h prior to the class containing information about the upcoming workshop, which contained their personalized primer text as preparation for the exercise. Students were unaware that more than one version of the scenario text existed.

In class, we asked students to review the information they had been given in the email and then to individually undertake a short observation on a video we provided. All students watched the same video, which was 62 s long and showed a flock of around 200 pigeons foraging on a patch of grass next to a road (Video S1). We asked students to first estimate the proportion of birds in the video that were actively foraging. This task was difficult and potentially subjective, given the very large number of individuals in frame and the small amount of time students were given to produce an estimate. We then asked students to select a specific pigeon in the flock and to observe it from the start to the end of the video, recording its total foraging (peck) rate for the duration of the observation. Students were then asked to conduct a second observation on a second, different pigeon of their choice. This task was less difficult and potentially less subjective. At the conclusion of the exercise, students entered their data into an online form. To confirm whether they had understood and remembered the primer, we also asked them whether, based on the information they had been given, they expected the pigeons to be hungry or satiated.

Supplementary video related to this article can be found at <https://doi.org/10.1016/j.anbehav.2024.03.013>

While students conducted other exercises as part of the class, we analysed the data and generated figures showing the pattern observed. At the end of the class, we revealed the hidden treatments, showed students the data, and invited them to discuss the results and their implications. We concluded the exercise by discussing ways in which potential bias could be reduced during data collection.

Ethical Note

Our experiment was approved by the University of Melbourne Human Ethics Team (reference number: 2023-27825-44541-2). This application confirms that all procedures meet the

requirements of the Australian National Statement on Ethical Conduct in Human Research, and that they were performed within the spirit of the Helsinki Declaration of 1975, as revised in 2000.

Statistical Analysis

We analysed our results using Bayesian generalized linear mixed models, fitted with the *brms* package (Bürkner, 2017) for R (version 4.3.1; R Core Team, 2023). To test whether an observer's a priori expectation affected estimated foraging proportion, we fitted two beta distributed models that each included a pair of fixed effects. We used two separate models because there were two different ways to code an observer's expectations. For the first model, we coded the fixed effect as an observer's allocated bias treatment. However, 24 of the 81 observers indicated a feeding motivation expectation opposite to that implied by the primer they were allocated. This suggested that the indicated expectation of an observer was likely to be a better predictor of observer bias than the allocated bias treatment. Hence, in a second model, we coded the fixed effect as an observer's indicated expectation. We also included cohort year as a fixed effect in the models. For both models, we found the difference between the posterior distributions of the median estimated foraging proportion by observers primed with the satiated example (or in the case of the second model those that believed pigeons were satiated) and those primed with the hungry example (or those that believed the pigeons to be hungry). We used the distribution of these differences to quantify the effect of the observer's a priori expectations. Following convention, we interpreted any differences with 95% credible intervals (CI) that did not overlap zero as biologically notable. For each model we specified a prior distribution of normal ($\mu = 0$; $\sigma = 1.5$) for fixed effects and exponential ($\lambda = 1$) for the dispersion parameter. These and all following priors are moderately informative; they aid model fit and make the models sceptical of extreme effect sizes.

To model the observer measured foraging rate of individual pigeons, we fitted a pair of negative binomial distributed models. These models followed the same structure (identical fixed effects and Bayesian modelling choices, see below) as the foraging proportion models, except that the response was an overdispersed count variable. We also included observer ID as a random effect to avoid pseudoreplication, as each observer measured the feeding rate of two pigeons. To test for an effect of bias treatment/a priori expectation, we calculated the difference contrast between posteriors as above. Prior distributions of normal ($\mu = 0$; $\sigma = 1.5$) for fixed effects and exponential ($\lambda = 1$) for the σ parameter were used in each model.

Allowing observers to haphazardly select a pigeon to observe may have provided an opportunity for selection bias. To test for this, we used a random number generator to select a subset of pigeons to estimate the 'true' mean feeding rate of the flock (Fig. A1). These data were inflated with zeros; we therefore fitted a zero-inflated Poisson model with an intercept term and a pigeon ID random effect (we counted the feeding rate for each pigeon twice) to model baseline feeding rate. We specified a prior distribution of normal ($\mu = 0$, $\sigma = 1.5$) for the intercept and exponential ($\lambda = 1$) for the σ parameter, the shape parameter and the zero-inflation parameter.

For all Bayesian models, we ran four chains each with 6000 iterations, preceded by 2000 warm-up iterations. A detailed description of our analysis can be viewed at https://tomkeaney.github.io/Biased_pigeons/.

RESULTS

Of the 178 students enrolled in the subject across the 3 years of the experiment, 82 provided 89 responses. Four students from the 2021 cohort provided multiple responses; we chose to only include

their initial response as we could not be sure that later responses were made while the students were still unaware of the experiment. One additional response did not provide the indicated expectation of the student, so it was removed from the data set. The cleaned data set therefore contained 81 responses from 81 students, 24 of which (29.6%) indicated an a priori expectation that did not match their allocated priming statement.

We found evidence of observer bias affecting measurement of pigeon foraging behaviour. However, the presence of this bias was contingent upon the trait observers were asked to score, as well as how a priori expectations were measured. There was no evidence that observer bias affected the estimated percentage of pigeons foraging: the difference between observers primed with a 'hungry' pigeon scenario and those primed with a 'satiated' pigeon scenario was just -0.52 percentage points (95% CIs: -9.20 to 8.43 ; Fig. 1a, b, Table A1). When the indicated expectation of the observer was modelled directly, those that believed the pigeons would be hungry estimated a foraging percentage that was 4.75 percentage points higher than those with a prior expectation of satiated pigeons, but this effect was unlikely to be biologically notable (95% CIs: -4.28 to 13.70 ; Fig. 1c, d, Table A2).

In contrast to foraging percentage, we found a large effect of observer bias on feeding rate observations. Several results demonstrate this finding. First, students massively overestimated the flock-wide mean feeding rate (Fig. 1e, g, Tables A3–A5). This was in large part caused by nonrandom selection of focal pigeons: only 16 of the 162 student observations were made on pigeons that did not feed (9.88%), whereas 12 of the 35 randomly selected pigeons did not forage at all throughout the entire video (34.3%). Second, the degree of overestimation was dependent upon the a priori expectation of the students. Those that expected hungry pigeons observed a feeding rate of 11.44 pecks/min (95% CIs: 8.05 to 15.94), compared to our baseline observation of 1.32 pecks/min (95% CIs: 0.54 to 2.53). In contrast, students that expected satiated pigeons observed a feeding rate 4.12 pecks/min lower than those that expected hungry pigeons (95% CIs: 0.938 to 8.04), equivalent to a 36.9% decrease (95% CIs: 9.8 to 56% ; Fig. 1g, h). However, this effect could not be detected if a priori expectation was modelled using the allocated bias treatment (difference = 1.58 pecks/min, 95% CIs: -1.96 to 5.35 , or a 15.8% decrease, 95% CIs: -22.6 to 41.8% ; Fig. 1e, f).

DISCUSSION

Our results suggest that the a priori expectations of observers affected their ability to accurately collect data. Observers that believed pigeons to be hungry estimated higher degrees of foraging than observers that believed pigeons to be satiated. However, observer bias was only detectable when expectations were modelled as the indicated expectation of the observer, rather than by the allocated primer we designed to influence the expectations of each observer. Unexpectedly, we also found that the effect size of the bias was greater when observers were asked to estimate the number of feeding events observed for a single pigeon, compared with when they were asked to estimate the proportion of the entire flock that was foraging. Below, we highlight some of the limitations of our experimental design, suggest improvements and highlight the broader implications our results have for animal behaviour research.

We designed our experiment with the putative expectation that foraging proportion was a more subjective trait than feeding rate. Subjective estimation is hypothesized to facilitate experimental bias, as interpretation is required to discriminate between categories. However, we found that the supposedly objective trait in our study, feeding rate, was overestimated to a greater extent by

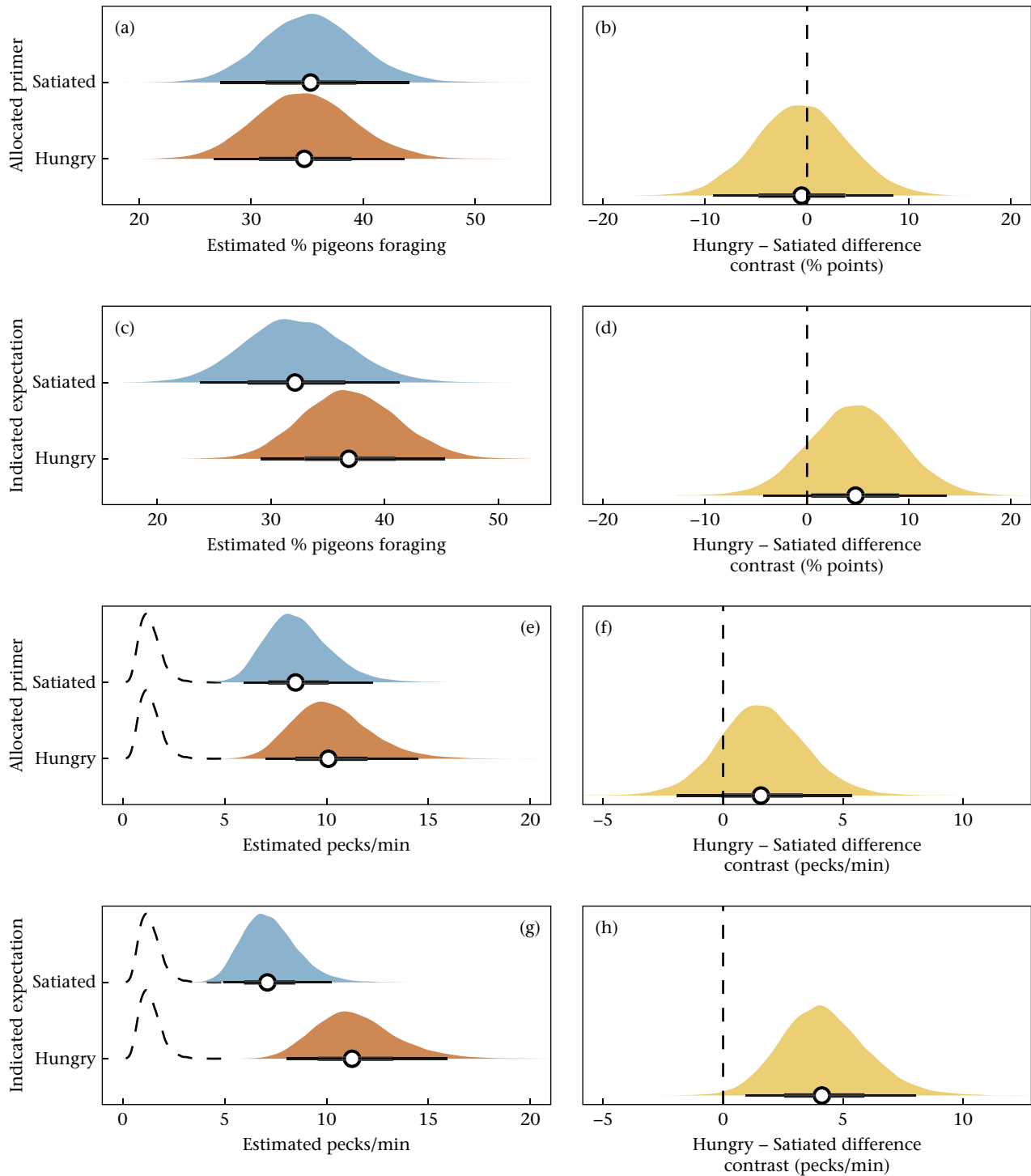


Figure 1. Posterior mean estimates and difference contrasts for (a–d) observer measured group foraging percentage and (e–h) individual feeding rates, split by whether we modelled the a priori beliefs of observers according to the priming statement they received or as the indicated expectation of the observer. The coloured area is the posterior distribution and the white point is the mean estimate with associated 67% and 95% credible intervals. The distributions shown with dashed lines are the posterior for baseline peck rate, estimated from a random sample of pigeons appearing in the video. Vertical dashed lines in (b) (d) (f) and (h) indicate a difference of zero between treatment groups.

observers that expected the pigeons to be hungry. A potential explanation for this surprising result is that while counting feeding attempts (peck rate) should be relatively objective, allowing observers to choose which pigeons to observe from the flock without restrictions or guidelines potentially introduced considerable subjectivity. Indeed, this selection bias was strong irrespective of a priori expectations, with pigeons that fed at least once being

chosen at a far higher rate than they were represented in the video footage. Thus, it is likely that observer bias was strong during estimation of this behaviour because the indicated expectation of observers elicited an effect on estimated feeding rate by strengthening the selection of feeding pigeons, rather than by increasing the overcounting of individual feeding attempts. Behavioural ecologists should think carefully about the data they collect and attempt to

make response variables as objective as possible. However, we encourage researchers to interpret this result as a cautionary note. Response variables that one considers to be objective are often likely to have some associated subjective elements that may not be obvious prior to data collection. Therefore, considerations of bias minimization, ideally through blinding, should be applied to all aspects of data collection.

A second unexpected finding from this study was that the expectations indicated by observers proved to better capture their a priori beliefs than did our ‘primed’ treatment groups. This was because the priming statements frequently failed to influence the expectations of observers in the expected direction: 24 of 81 observers indicated an expectation opposite to what their priming statement suggested. The primers were explicit about the condition of the pigeons, suggesting that these observers simply did not read them, students discussed their priming statements prior to the experiment or that they were difficult to comprehend. The first explanation seems most likely, for several reasons. First, the primer texts were short, in plain English and did not contain jargon. Second, students were unaware of the importance of the priming text for the experiment, and not informed that other participants had received different primers. While student discussion of the email was possible, no student in any cohort indicated that they were aware of the experiment, even when discussing the results during the final stage of the workshop. The concealed importance of the priming information provided the benefit that observers did not consciously try to avoid observer bias. However, as a consequence it was difficult to impress the importance upon observers of carefully reading the priming statement and keeping this knowledge confidential. Importantly, these failures during the priming process would have masked the effect of observer bias we aimed to detect. Only by modelling expectations as those directly indicated by the observers were we able to detect observer bias. A second piece of cautionary advice is thus required: those replicating our experiment in the future should think carefully about the design of their priming treatment, with an emphasis on delivering it through an effective medium.

As highlighted above, our experimental design has limitations. For those interested in replicating our experiment, we suggest an improved design (Fig. 2). Students can be primed in similar fashion the day before the experiment, with the addition of a request for a read receipt where students should indicate their expectation on pigeon hunger levels. This will allow researchers to ascertain the proportion of students that were appropriately primed prior to the experiment. Measures could also be taken to reduce the risk of students discussing the emails they received, for example by asking students to keep their priming statements confidential. During the class students can then estimate the proportion of the pigeons feeding during the filmed interval as presented in this paper. However, when observing the feeding rate of specific pigeons, we recommend the use of random number generation combined with a grid to select pigeons for observation (see Fig. A1 for an example). This shifts the focus of data collection onto peck rate explicitly, rather than the selection of the individual in the first instance. The design change removes the large subjective component of this response variable, which we did not a priori predict when designing our experiment. With these changes, the experiment should prime students effectively and test how their a priori expectations affect measurement of foraging behaviours that are truly at different ends of the subjectivity – objectivity spectrum.

Previous studies have suggested that undergraduate students are not representative of active researchers, as they lack observational experience (Tuytens et al., 2014; van Wilgenburg & Elgar, 2013). Observer bias may therefore not be the problem our evidence, and that from related studies, suggests. However, while undergraduates may face greater difficulty discriminating between similar behaviours (potentially making estimation more subjective), they are not incentivized to produce estimates that support their a priori expectations to the same extent as currently practising researchers (Marsh & Hanlon, 2004). In stark comparison with the publishing and downstream funding benefits researchers gain from producing ‘significant’ results (Fanelli, 2012; Grimes et al., 2018), in the studies that demonstrated observer bias using undergraduate cohorts, hypothesis testing has not been used as part of the design

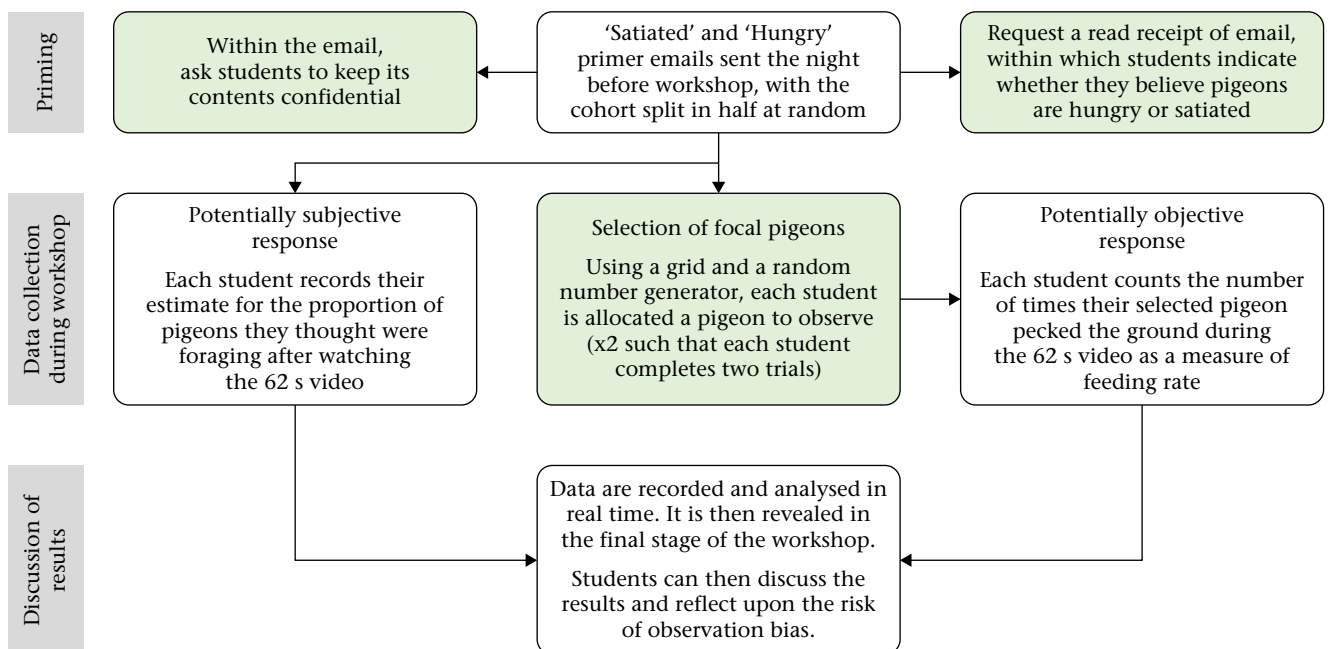


Figure 2. Schematic for improved experimental design that could be used to replicate the experiment. Boxes in green show improvements that were not included in the present study.

(Marsh & Hanlon, 2004; Rosenthal & Fode, 1963; Tuytens et al., 2014; present study). Furthermore, there is a wealth of evidence that biases affect this field (Fraser et al., 2018). In addition to the indirect evidence for observer bias provided by comparing blind with nonblind studies (Holman et al., 2015; van Wilgenburg & Elgar, 2013), p hacking (Head et al., 2015), publication bias (Jennions et al., 2013; Yang et al., 2023) and reviewer bias (Fox et al., 2023) have all been robustly demonstrated. We therefore believe that undergraduate students being more prone to observer bias is an unjustified claim. The evidence instead suggests that our results should if anything be taken as a conservative estimate for what occurs in the behavioural ecology field at large.

Our experiment demonstrates how expectation bias can subconsciously influence the process of data collection in a cohort of university students nearing graduation. In discussions at the conclusion of the class, students noted that the revelation of the hidden treatment and the results of the experiment provided them with a more vivid and memorable demonstration of their own potential for bias than alternative teaching approaches such as a class discussion of sources of bias might have. We encourage others to explore ways of teaching students about the important issue of bias, and we offer our approach as one such potentially effective way. We welcome the use of the materials we have developed for undergraduate teaching and provide them freely in the [Appendix](#). Finally, we highlight that a fallibility in our experimental design, that our supposedly objective response variable actually contained a strong subjective component that we did not predict prior to data collection, helps provide an answer as to when behavioural ecologists should employ bias mitigating strategies such as blind data recording. That is, they should do so whenever possible, as it is very difficult to judge how robust any one response variable is against observation bias.

Author Contributions

Thomas Keaney: Methodology, Formal analysis, Investigation, Data curation, Writing – Original draft, Writing – review and editing, Visualization; **Theresa Jones:** Conceptualization, Methodology, Investigation, Writing – Original draft, Writing – review and editing; **Raoul Mulder:** Conceptualization, Methodology, Investigation, Writing – Original draft, Writing – review and editing, Project administration.

Data Availability

Teaching materials are available in the Teaching_resources folder found in the github repository https://github.com/tomkeaney/Biased_pigeons. For a detailed report of our analysis, see https://tomkeaney.github.io/Biased_pigeons/. The report contains all raw data and R-script required to reproduce our analysis, build [Fig. 1](#) and produce the Appendix Tables. Additionally, our raw data are deposited in the Open Science Framework database: <https://osf.io/zmb7k/>.

Declaration of Interest

We have no conflicts of interest to declare.

Acknowledgments

We are grateful to the students that enthusiastically participated in the Animal Behaviour course at the University of Melbourne. We also thank Guru Guillera-Arroita and José Lahoz-Monfort for filming the pigeon video, and Annie Aulsebrook and Marty Lockett for assistance with data collection.

References

- Balph, D. F., & Balph, M. H. (1983). On the psychology of watching birds: The problem of observer-expectancy bias. *Auk*, *100*(3), 755–757.
- Bello, S., Krogsbøll, L. T., Gruber, J., Zhao, Z. J., Fischer, D., & Hróbjartsson, A. (2014). Lack of blinding of outcome assessors in animal model experiments implies risk of observer bias. *Journal of Clinical Epidemiology*, *67*(9), 973–983.
- Bertram, B. C. R. (1980). Vigilance and group size in ostriches. *Animal Behaviour*, *28*(1), 278–286.
- Burghardt, G. M., Bartmess-LeVasseur, J. N., Browning, S. A., Morrison, K. E., Stec, C. L., Zachau, C. E., & Freeberg, T. M. (2012). Perspectives – Minimizing observer bias in behavioral studies: A review and recommendations. *Ethology: Formerly Zeitschrift Fur Tierpsychologie*, *118*(6), 511–517.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1.
- Crossley, N. A., Sena, E., Goehler, J., Horn, J., van der Worp, B., Bath, P. M. W., Macleod, M., & Dirnagl, U. (2008). Empirical evidence of bias in the design of experimental stroke studies: A metaepidemiologic approach. *Stroke; a Journal of Cerebral Circulation*, *39*(3), 929–934.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904.
- Fox, C. W., Meyer, J., & Aimé, E. (2023). Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology*, *37*(5), 1144–1157.
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS One*, *13*(7), Article e0200303.
- Grimes, D. R., Bauch, C. T., & Ioannidis, J. P. A. (2018). Modelling science trustworthiness under publish or perish pressure. *Royal Society Open Science*, *5*(1), Article 171511.
- Hannah, S. D., & Carpenter-Song, E. (2013). Patrolling your blind spots: Introspection and public catharsis in a medical school faculty development course to reduce unconscious bias in medicine. *Culture, Medicine and Psychiatry*, *37*(2), 314–339.
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*(3), Article e1002106.
- Hirst, J. A., Howick, J., Aronson, J. K., Roberts, N., Perera, R., Koshiaris, C., & Heneghan, C. (2014). The need for randomization in animal trials: An overview of systematic reviews. *PLoS One*, *9*(6), Article e98856.
- Holman, L., Head, M. L., Lanfear, R., & Jennions, M. D. (2015). Evidence of experimental bias in the life sciences: Why we need blind data recording. *PLoS Biology*, *13*(7), e1002190–e1002190.
- Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., Ravaut, P., & Brorson, S. (2012). Observer bias in randomised clinical trials with binary outcomes: Systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*, *344*, Article e1119.
- Jennions, M. D., Lortie, C. J., Rosenber, M. S., & Rothstein, H. R. (2013). Publication and related biases. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of meta-analysis in ecology and evolution* (pp. 207–236). Princeton University Press.
- Kaptchuk, T. J. (2003). Effect of interpretive bias on research evidence. *BMJ*, *326*(7404), 1453–1455.
- Kardish, M. R., Mueller, U. G., Amador-Vargas, S., Dietrich, E. I., Ma, R., Barrett, B., & Fang, C.-C. (2015). Blind trust in unblinded observation in ecology, evolution, and behavior. *Frontiers in Ecology and Evolution*, *3*, Article 142945.
- Kozlov, M. V., Zverev, V., & Zvereva, E. L. (2014). Confirmation bias leads to over-estimation of losses of woody plant foliage to insect herbivores in tropical regions. *PeerJ*, *2*, Article e709.
- MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*, 259–287.
- Marsh, D. M., & Hanlon, T. J. (2004). Observer gender and observation bias in animal behaviour research: Experimental tests with red-backed salamanders. *Animal Behaviour*, *68*(6), 1425–1433.
- Mincey, K. D. (2021). Teaching bias to undergraduate students in an introduction to public health course. *Pedagogy in Health Promotion*, *7*(4), 313–316.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, *2*(2), 175–220.
- R Core Team. (2023). R: A language and environment for statistical computing. <https://www.R-project.org/>.
- Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Systems Research: The Official Journal of the International Federation for Systems Research*, *8*(3), 183–189.
- Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA: The Journal of the American Medical Association*, *273*(5), 408–412.
- Schulz, K. F., & Grimes, D. A. (2002). Blinding in randomised trials: Hiding who got what. *Lancet*, *359*(9307), 696–700.
- Traniello, J. F. A., & Bakker, T. C. M. (2015). Minimizing observer bias in behavioral research: Blinded methods reporting requirements for behavioral ecology and sociobiology. *Behavioral Ecology and Sociobiology*, *69*(10), 1573–1574.
- Tuytens, F. A. M., de Graaf, S., Heerkens, J. L. T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E., & Ampe, B. (2014). Observer bias in animal behaviour research: Can

- we believe what we score, if we score what we believe? *Animal Behaviour*, 90, 273–280.
- Tuytens, F. A. M., Stadig, L., Heerkens, J. L. T., Van laer, E., Buijs, S., & Ampe, B. (2016). Opinion of applied ethologists on expectation bias, blinding observers and other debiasing techniques. *Applied Animal Behaviour Science*, 181, 27–33.
- Van Deventer, A., & Shrader, A. M. (2021). Predation risk and herd position influence the proportional use of antipredator and social vigilance by impala. *Animal Behaviour*, 172, 9–16.
- van Wilgenburg, E., & Elgar, M. A. (2013). Confirmation bias in studies of nestmate recognition: A cautionary note for research into the behaviour of animals. *PLoS One*, 8(1), Article e53548.
- Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., Gluud, C., Martin, R. M., Wood, A. J. G., & Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *BMJ*, 336(7644), 601–605.
- Yang, Y., Sánchez-Tójar, A., O’Dea, R. E., Noble, D. W. A., Koricheva, J., Jennions, M. D., Parker, T. H., Lagisz, M., & Nakagawa, S. (2023). Publication bias impacts on effect size, statistical power, and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology. *BMC Biology*, 21(1), 71.
- Zvereva, E. L., & Kozlov, M. V. (2019). Biases in studies of spatial patterns in insect herbivory. *Ecological Monographs*, 89(3), e01361.

Appendix

Priming emails

‘Hungry forager’ primer:

Dear [student name]

As part of the Skills Studio tomorrow, we would like you to conduct a short observation. It will take about 5 minutes to complete.

You will complete the observation during the class – this email is just to provide you with the links in advance.

If you can’t come to the class, please complete the observation anyway (before the start of the class).

About the exercise

In this exercise, we will transport you to a location in Brunswick where a flock of pigeons is feeding.

The video was recorded just after winter, which is typically the most difficult season of the year for pigeons in terms of feeding and survival. About 30% of pigeons do not survive winter; many that do survive experience a significant loss of body mass and must make the most of every feeding opportunity that arises in order to regain condition.

Your task

Access the 1-minute video of feeding pigeons at: [video link]

Before you start playing the video, click on the ‘cog’ (Source Chooser) icon and select the highest resolution (2628 kbps).

Start playing the video in full screen mode

1. Scan around the flock and try to estimate the percentage of pigeons that are foraging. There is a lot going on, so this is challenging, but try to come up with a figure that feels reasonably accurate to you. Write down this percentage.
2. Pick out a pigeon in the flock to do a focal observation on. Restart the video and count how many times this bird pecks the ground during the 1-minute

(continued)

video. Write down your count. Now, pick another bird and go back to the start of the video again. Also record the peck count for this second bird.

Click on this link to enter your data: [survey link]

That’s it – you’re done!

Thank you! [course coordinator names]

‘Satiated forager’ primer:

Dear [student name]

As part of the Skills Studio tomorrow, we would like you to conduct a short observation. It will take about 5 minutes to complete.

You will complete the observation during the class – this email is just to provide you with the links in advance.

If you can’t come to the class, please complete the observation anyway (before the start of the class).

About the exercise

In this exercise, we will transport you to a location in Brunswick where a flock of pigeons is feeding.

The video was recorded late in the afternoon near Jewell railway station, a favourite spot for commuters to feed the pigeons. Moreland Council estimates that locals dump 5–10 kg of bread and seed here, creating a superabundant daily food source. By the end of the day, many of the birds are completely satiated and have little or no motivation to feed.

Your task

Access the 1-minute video of feeding pigeons at: [video link]

Before you start playing the video, click on the ‘cog’ (Source Chooser) icon and select the highest resolution (2628 kbps).

Start playing the video in full screen mode

1. Scan around the flock and try to estimate the percentage of pigeons that are foraging. There is a lot going on, so this is challenging, but try to come up with a figure that feels reasonably accurate to you. Write down this percentage.

2. Pick out a pigeon in the flock to do a focal observation on. Restart the video and count how many times this bird pecks the ground during the 1-minute video. Write down your count. Now, pick another bird and go back to the start of the video again. Also record the peck count for this second bird.

Click on this link to enter your data: [survey link]

That’s it – you’re done!

Thank you! [course coordinator names]

Estimating baseline feeding rate

We estimated a baseline peck rate by observing 35 pigeons. To select the pigeons we observed, we split a still image of the foraging video (taken at time zero) into a 43 × 21 cell grid. From the 122 cells that contained pigeons, 40 were chosen by random number generation. If multiple pigeons were present in the cell, we selected the most prominent to observe. Five observations were discarded: three due to overlap of the same pigeon between cells that were selected by the random number generator and two more as the pigeons left the field of view during the video and could no longer be tracked.

Table A1

Posterior estimates of the percentage of pigeons foraging, split by the primer that students were allocated

Bias treatment	No. of students	Estimated % foraging	Estimated error	2.5% CI	97.5% CI
Hungry	44	34.84	4.32	26.7	43.68
Satiated	37	35.36	4.27	27.24	44.09

CI = credible interval.

Table A2

Posterior estimates of the percentage of pigeons foraging, split by the actual expectation of the students

Indicated expectation	No. of students	Estimated % foraging	Estimated error	2.5% CI	97.5% CI
Hungry	45	36.98	4.17	29.13	45.32
Satiated	36	32.24	4.49	23.75	41.32

CI = credible interval.

Table A3

The estimated peck rate of foraging pigeons, split by the primer students were allocated

Bias treatment	No. of pigeons observed	Estimated peck rate	Estimated error	2.5% CI	97.5% CI
Hungry	88	10.25	1.91	7.02	14.48
Satiated	74	8.64	1.62	5.94	12.26
Baseline	35	1.32	0.51	0.54	2.53

Each student observed two pigeons. CI = credible interval.

Table A4

The estimated peck rate of foraging pigeons, split by the indicated expectation of the students

Indicated expectation	No. of pigeons observed	Estimated peck rate	Estimated error	2.5% CI	97.5% CI
Hungry	90	11.44	2.01	8.05	15.94
Satiated	72	7.23	1.36	4.94	10.25
Baseline	35	1.32	0.51	0.54	2.53

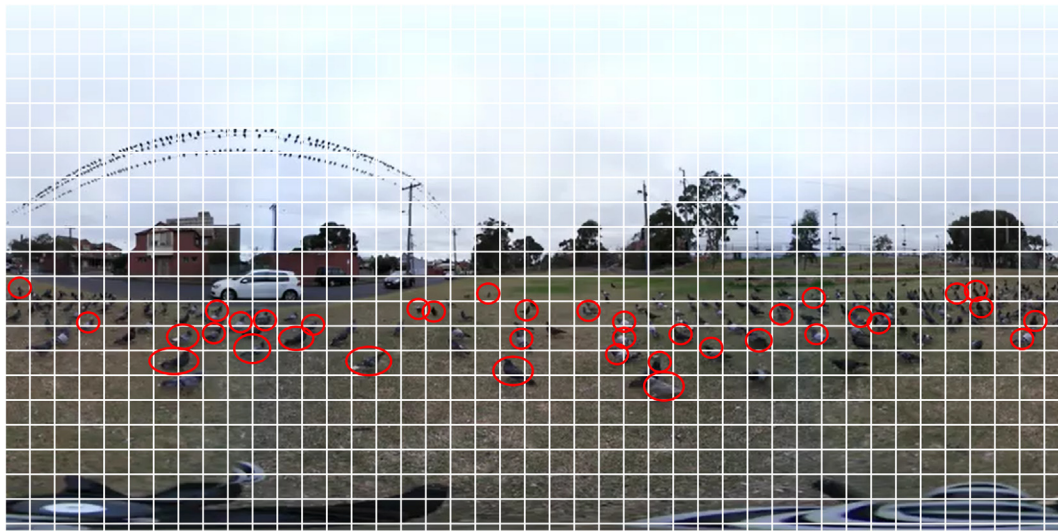
Each student observed two pigeons. CI = credible interval.

Table A5

The degree to which each group of students overestimated feeding rate (number of ground pecks/min)

Statistic	Median	Estimated error	2.5% CI	97.5% CI
Bias treatment Hungry/Baseline	7.83	4.34	3.53	19.61
Bias treatment Satiated/Baseline	6.58	3.65	2.97	16.49
Expectation Hungry/Baseline	8.74	4.82	4.01	22.16
Expectation Satiated/Baseline	5.5	3.1	2.47	13.92

CI = credible interval.

**Figure A1.** The pigeons selected for baseline estimation of feeding rate.