Replication in behavioural ecology: a comment on Ihle et al.

Ben J. Hatchwell[a]
[a]Department of animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

Ihle et al. (2017) have made a valuable contribution to a current debate on open science. The general sentiment that scientific disciplines should be reliable, reproducible and replicable should, of course, hold for behavioural ecology as much as it does for any other scientific discipline, and while the extent of some of the problems they describe in our field is a matter for debate, few would disagree that we should adopt practices that enhance the transparency and credibility of behavioural ecology.

One issue arising from Ihle et al.'s (2017) prospectus concerns their suggestions regarding replication. Behavioural ecologists study evolutionary adaptations of organisms to their environment, and variation is an inherent property of that dynamic process. This variation is observable at every level of biological organization, but we are particularly concerned by the causes and consequences of variation among individuals, populations and species. So, how replicable should we expect studies to be, and how do we interpret failure to replicate? For example, in studies of cooperatively breeding birds comparisons across populations have revealed drivers of cooperation and led to novel insights on benefits of helping (e.g. Koenig and Stacey 1990; Reyer 1990; Baglione et al. 2002; Sharp et al. 2011) because of population-level differences. Likewise, long-term studies of single populations have shown that while some effects may be stable, temporal variation in ecology or social environment may drive marked changes in helping behaviour (Koenig and Dickinson 2016). Thus, replicated analyses across populations or through time cannot be treated as validatory exercises. Inherent variation may be better understood and controlled in model systems, although even here replicated studies often produce inconsistent results in lab (Seguin and Forstmeier 2012) and field (Parker 2013), the basis for which may be uncertain. Therefore, while there is a strong argument to be made for greater replication, the motive for doing so should not necessarily be one of validation, as implied by Ihle et al. (2017). The strong theoretical basis to behavioural ecology makes it a robust, predictive science, but we should not lose sight of the fact that the dynamic response of organisms to their environment also makes it an essentially variable and often intriguingly complex one.

The other difficulty with replication is that virtually every new research programme, whether a grant application, PhD studentship proposal or new line of enquiry in an established study, starts with the question: is it novel or original and will it produce outputs that have some impact and advance the field? Rightly or wrongly, the incentive of funding agencies to direct resources towards replication is low, and for individual scientists (especially early career researchers trying to build a reputation for cutting-edge work) there may be little benefit from replicating previous studies; ironically, this may be particularly true if the same conclusion is drawn. This is not to deny the desirability of replication, but simply recognizes the realities of funding, career advancement and editorial practice.

The other recommendations of Ihle et al. (2017) largely formalize what should be regarded as good scientific practice, e.g. "blinding" of experiments (even though this may not always be practical, especially in field studies), testing of "a priori" hypotheses, etc. The key change they advocate is in the transparency of the processes involved in the conduct of research; one could characterize this agenda as one of opening for scrutiny the lab and field notebooks in which are recorded the hypotheses we test, the predictions we make, the data we collect and use, and the analyses and interpretation of our results. One potential benefit of this would be to improve the connection between data and results. I am probably not alone in sometimes being frustrated by the difficulty of relating the outcomes of statistical models in publications to the data they describe—the formal documenting of workflow and process of analysis would certainly increase transparency in this regard. Of course, this begs the question of where the onus lies to scrutinize preregistrations and workflows, and raises the concern that reviewer goodwill may be exhausted if their task is made more onerous by increasing journals' expectations of them. This will be a question at the forefront of editors' minds as moves to increase credibility in all scientific disciplines develop.

Address correspondence to B.J. Hatchwell. E-mail: b.hatchwell@sheffield.ac.uk

Received 13 December 2016; editorial decision 23 December 2016; accepted 10 January 2017.

doi:10.1093/beheco/arx009

Editor-in-chief: Leigh Simmons

REFERENCES

Baglione V, Canestrari D, Marcos JM, Griesser M, Ekman J. 2002. History, environment and social behaviour: experimentally induced cooperative breeding in the carrion crow. Proc R Soc Lon B 269:1247–1251.
Ihle M, Winney IS, Krysalli, Croucher M. 2017. Striving for transparent and credible research: practical guidelines for behavioral ecologists. Behav Ecol. 28: 348–354.
Koenig WD, Dickinson JL. 2016. Cooperative breeding in vertebrates: studies of ecology, evolution, and behavior. Cambridge (UK): Cambridge University Press.
Koenig WD, Stacey PB. 1990. Group-living and food storage under contrasting ecological conditions. In: Stacey PB, Koenig WD, editors. Cooperative breeding in birds: long-term studies of ecology and behaviour. Cambridge (UK): Cambridge University Press. p. 413–453.
Parker TH. 2013. What do we really know about the signaling role of plumage colour in blue tits? A case study of impediments to progress in evolutionary biology. Biol Rev. 88:511–536.
Reyer H-U. 1990. Pied kingfishers: Ecological causes and reproductive consequences of cooperative breeding. In: Stacey PB, Koenig WD, editors. Cooperative breeding in birds: long-term studies of ecology and behavior. Cambridge (UK): Cambridge University Press. p. 529–557.
Seguin A, Forstmeier W. 2012. No band color effects on male courtship rate or body mass in the zebra finch: four experiments and a meta-analysis. PLoS One. 7:e37785.
Sharp SP, Simeoni M, McGowan A, Nam KB, Hatchwell BJ. 2011. Patterns of recruitment, relatedness and cooperative breeding in two populations of long-tailed tits. Anim Behav. 81:843–849.

Transparent and credible practices under the microscope: a response to comments on Ihle et al.

Isabel S. Winney[a] and Malika Ihle[b]
[a]Laboratoire Evolution & Diversité Biologique, Bâtiment 4R1, Université de Toulouse Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France and [b]Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

Improving research practices is a community project, and as such we are excited to have received many responses to our paper (Ihle et al. 2017). Overall, preregistration provoked the most reactions,

ranging from enthusiasm to reluctance to adopt the measures, and we will dedicate most of our response to this topic. The remaining comments focused on replications. Hatchwell (2017) highlighted that they provide the raw data to study the underlying cause of variation. Almost in answer to this, Parker and Nakagawa (2017) proposed that journals append each replication to the original study, and that researchers launch groups of related replications or "replication batteries," so that we can begin to understand the variation in our systems. Before addressing some common misunderstandings about preregistration, we need to stress that, as publically funded scientists, we are accountable for the work that we do. Facing drastic funding cuts and the erosion of public confidence, this is not the time to argue about whether unreliable, irreproducible, and nonreplicable methods are more fun or bring us personal success; it is the time to embrace more rigorous scientific practices.

Preregistration was the most contentious proposal, with Koenig (2017) arguing that it stifles creativity and is unnecessary due to low type 1 error rates, Cockburn (2017) voicing doubts over preregistering studies that may change over time, and Blumstein (2017) expressing concerns about being scooped. We reiterate that the more freedom individuals have to reshape analyses, the higher the type 1 error rate climbs (Forstmeier et al. 2016). Consequently, the number of false positives from exploratory analyses will be higher than suggested by Koenig (2017), and it can only be reduced by documenting a priori what tests will be performed. Preregistration involves good planning, taking the creative time that previously would have been distributed throughout a project and re-allocating it to the start. Ultimately, this allows researchers to prevent mistakes and streamlines both analysis and writing (see Forstmeier 2017), which we find satisfying and rewarding. Preregistrations are not the end of the creative process: they are merely the part that we can guarantee is relatively free of bias. Exploratory analyses *can still be published*, right alongside preregistered analyses. They are an important tool for generating new hypotheses, but they should be distinguished because they are, by definition, influenced by viewing the data. Lastly, researchers do not need to worry about protecting their ideas: preregistrations can be embargoed for up to 4 years on the Open Science Framework or as long as desired on AsPredicted.org. Besides, opening up preregistration (or part of it) right at conception could promote collaboration, mirroring the approach taken in epidemiology (Cockburn 2017).

With preregistration and replication, we can at last reduce the likelihood and mitigate the impact of type 1 errors and get on with the business of understanding variation. Writing preregistrations and replication reports, as well as protocols and codes for data preparation and data analyses, will improve the quality of scientific work, while opening up this documentation will allow researchers to get credit for the work they actually do. We encourage everyone to embrace measures that we know increase scientific rigor (e.g., Kidwell et al. 2016), and ask the unwilling not to impede their spread while others discover their benefits.

Eventually, "we must choose between what is easy and what is right" (to quote Dumbledore in *Harry Potter and the Goblet of Fire* [Rowling 2000]). As long as our goal is to aim to get closer to the truth, we know that improving our practices is part of our fundamental remit as scientists. Now that multiple tools and protocols have been developed to make that process easier, all that is left is for you to use them.

FUNDING

M.I. was funded by a grant from the Natural Environment Research Council (NERC, MO 005941) awarded to T. Burke. I.S.W. was funded by an ERC grant to B. Pujol (ANGI 681484).

We thank all attendees of the Life History and Behaviour journal club at the University of Sheffield and the Laboratoire Evolution & Diversité Biologique journal club at the University Paul Sabatier, Toulouse, for giving us their feedback regarding the initial invited idea. We also thank A. Krystalli for discussions, and A. Krystalli, J. Pick, W. Forstmeier, and T. Parker for comments on the final version of the manuscript.

Address correspondence to I.S. Winney. E-mail: isabel.winney@gmail.com.

Received 26 January 2017; revised 30 January 2017; editorial decision 31 January 2017; accepted 3 March 2017.

doi:10.1093/beheco/arx026

Editor-in-Chief: Leigh Simmons

REFERENCES

Blumstein DT. 2017. Research credibility: the devil is in the details: a comment on Ihle et al. Behav Ecol. 28:355.

Cockburn A. 2017. Long-term data as infrastructure: a comment on Ihle et al. Behav Ecol. 28: 357.

Forstmeier W. 2017. Preregister now for an upgrade to Behavioral Ecology 2.0: a comment on Ihle et al. Behav Ecol. 28: 358–359.

Forstmeier W, Wagenmakers E, Parker TH. 2016. Detecting and avoiding likely false-positive findings - a practical guide. Biol Rev. doi:10.1111/brv.12315.

Hatchwell BJ. 2017. Replication in behavioural ecology: a comment on Ihle et al. Behav Ecol. 28: 360.

Ihle M, Winney IS, Krystalli A, Croucher M. 2017. Striving for transparent and credible research: practical guidelines for behavioral ecologists. Behav Ecol. 28: 348–354.

Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg LS, Kennett C, Slowik A, Sonnleitner C, Hess-Holden C, et al. 2016. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. PLoS Biol. 14:e1002456.

Koenig W. 2017. Striving for science that is transparent, credible—and enjoyable: a comment on Ihle et al. Behav Ecol. 28: 358.

Parker TH, Nakagawa S. 2017. Practical models for publishing replications in behavioral ecology: a comment on Ihle et al. Behav Ecol. 28: 355–357.

Rowling JK. 2000. Harry Potter and the Goblet of Fire. London (UK): Bloomsbury.