BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Comment

Cite this article: Durkee PK. 2019 Do the Maasai perceive weak walkers to be stronger and more attractive than strong walkers? A re-analysis of Fink *et al*. (2019). *Biol. Lett.* **15**: 20190240.
http://dx.doi.org/10.1098/rsbl.2019.0240

Received: 1 April 2019
Accepted: 13 May 2019

Subject Areas:
behaviour, cognition, evolution

Author for correspondence:
Patrick K. Durkee
e-mail: pdurkee@utexas.edu

The accompanying reply can be viewed at http://dx.doi.org/10.1098/rsbl.2019.0376.

THE ROYAL SOCIETY PUBLISHING

Evolutionary biology

Do the Maasai perceive weak walkers to be stronger and more attractive than strong walkers? A re-analysis of Fink *et al*. (2019)

Patrick K. Durkee

The University of Texas at Austin, 108 E Dean Keeton St, Austin, TX 78712, USA

PKD, 0000-0001-6159-4277

Fink *et al*. [1] report that Maasai men and women rated strong walkers as both weaker and less attractive than weak walkers: a surprising reversal of the effects found in previous research [2,3]. These data from a non-WEIRD population are highly valuable for assessing the universality gait perception [4]. However, the conducted analyses relied on participants' ratings of attractiveness and strength that were averaged across groups of stimuli (i.e. by-participant aggregation), which is problematic for two reasons (see [5–7] for a full discussion of these issues). First, aggregating compounds error variance and severely inflates false-positive rates. Second, inferences from by-participant aggregation can only be applied to those *exact stimuli* used in the study because averaging treats the stimuli as a fixed effect rather than a random sample from the population of potential stimuli. These problems are pervasive in the literature and not unique to Fink *et al*. [1], but given the current state of psychological research, it is critical to begin to examine and address these issues where they occur.

I downloaded the data presented in Fink *et al*. [1] and attempted to reproduce the reported effects using a mixed-modelling approach, which (a) accounts for the random variance between raters and targets that inflates false-positive rates, and (b) allows for generalization of inferences to other samples of both raters and targets by treating both as a random effect. I ran a separate ordinal mixed-effect model for each outcome (i.e. strength and attractiveness ratings) using the clmm function in R [8,9]. I used likelihood ratio tests to compare models with different random effects structures through a top-down approach: starting with the maximal random effects structure and reducing based on model fit.

For the strength perceptions outcome, the maximal model—which allows random intercepts for both targets and raters, and random slopes for raters (i.e. the effect of walker group can differ between raters)—provided somewhat better fit than a reduced model that only specified random intercepts for raters and targets ($p = 0.047$). In this model, the correlated error-variance for each target ($\sigma^2 = 1.18$) and rater ($\sigma^2 = 0.52$) is accounted for, and the surprising effect where strong walkers are perceived as less strong than weak walkers disappears ($b = 0.606$, $p = 0.263$). Thus, we cannot infer that 'strong' British walkers are perceived as either systematically weaker or stronger by Maasai men and women than 'weak' British walkers.

For the attractiveness-perceptions outcome, the maximal model did not provide significantly better fit than a reduced model specifying random intercepts but fixed slopes for both raters and targets ($p = 0.162$); this intercept-only was much better than a further reduced model that fixed target intercepts ($p < 0.001$). This suggests that there is no significant variation in the effects of walker-group on raters' perceptions of attractiveness and that allowing random slopes to vary would be an over-specification of the model [10,11];

Therefore, I chose to use the more parsimonious intercept-only model. In this model, which accounts for the random variance between targets ($\sigma^2 = 0.83$) and raters ($\sigma^2 = 0.53$), there was no difference in raters' perceptions of attractiveness between strong and weak walkers ($b = 0.366$, $p = 0.337$). Thus, we cannot infer that Maasai men and women systematically view 'strong' British walkers as either more or less attractive than 'weak' British walkers.

Fink *et al*. [1] also report a positive correlation between ratings of strength and attractiveness. This analysis also suffers from the issues of by-participant aggregation, so I examined this effect using the more appropriate mixed-modelling approach. There were no significant differences in model fit between models allowing random intercepts and slopes to vary between both raters and targets, and models that only allow intercepts to vary at each level ($p = 0.617$), so I chose the more parsimonious intercept-only model as the final model. I modelled the predictor (i.e. strength) as a linear variable rather than ordinal because it did not significantly worsen model fit to do so ($p = 0.976$). The results from this model reveal that the intra-individual Maasai ratings of attractiveness are reliably positively associated with ratings of strength ($b = 0.314$, $p < 0.001$), suggesting that Maasai perceptions of attractiveness do indeed track perceptions of strength.

Finally, I note some methodological issues present in Fink *et al*. [1] and offer solutions with the intention of improving this interesting research programme. The stimuli were created by dichotomizing from a continuous set of 70 viable videos, which reduces power and can lead to spurious findings [12,13]. Studies employing a design where two groups of 10 videos are shown to participants (e.g. [1–3]) *never* reach 80% power to detect a 'medium' sized effect ($d = 0.5$) even with an infinite number of raters [10]. If future studies use the targets' actual handgrip strength as a predictor of strength and attractiveness ratings in a mixed-model, they would reach 80% power to detect that effect with only 40 raters—and could detect smaller effects if all 70 existing walker stimuli were used [10]. Moreover, such an approach would be more consistent with the underlying research question.

Fink *et al*. [1] provided valuable data on an interesting topic, but methodological and analytical choices undermine the veracity of their findings. When random variance between targets is handled appropriately rather than aggregated, there is no difference in Maasai ratings of strength or attractiveness between 'weak' British walkers and 'strong' British walkers. This null effect is still theoretically important and interesting: it suggests that more research is needed to understand whether and how gait informs human person perception. I hope that future studies on the topic will fruitfully employ the more rigorous methods and robust analyses discussed here.

Data accessibility. The data used in this re-analysis are available from the original authors on figshare (https://doi.org/10.6084/m9.figshare.7334447). The R-scripts used to conduct the analysis and produce the manuscript are available at on the OSF (https://osf.io/kcx5j/?view_only=0e1adbb732b64c53a37e4df702642295).

Competing Interests. I declare I have no competing interests.

Funding. I received no funding for this study.

References

1. Fink B, Butovskaya ML, Shackelford TK. 2019 Assessment of physical strength from gait: data from the Maasai of Tanzania. *Biol. Lett.* **15**, 20180803. (doi:10.1098/rsbl.2018.0803)
2. Fink B, André S, Mines JS, Weege B, Shackelford TK, Butovskaya ML. 2016 Sex difference in attractiveness perceptions of strong and weak male walkers. *Am. J. Hum. Biol.* **28**, 913–917. (doi:10.1002/ajhb.22891)
3. Fink B, Wübker M, Ostner J, Butovskaya ML, Mezentseva A, Muñoz-Reyes JA, Sela Y, Shackelford TK. 2017 Cross-cultural investigation of male gait perception in relation to physical strength and speed. *Front. Psychol.* **8**, 1427. (doi:10.3389/fpsyg.2017.01427)
4. Gurven MD. 2018 Broadening horizons: sample diversity and socioecological theory are essential to the future of psychological science. *Proc. Natl Acad. Sci. USA* **115**, 11 420–11 427. (doi:10.1073/pnas.1720433115)
5. Clark HH. 1973 The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. Verbal Learn. Verbal Behav.* **12**, 335–359. (doi:10.1016/S0022-5371(73)80014-3)
6. Fontenelle GA, Phillips AP, Lane DM. 1985 Generalizing across stimuli as well as subjects: a neglected aspect of external validity. *J. Appl. Psychol.* **70**, 101–107. (doi:10.1037/0021-9010.70.1.101)
7. Judd CM, Westfall J, Kenny DA. 2012 Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *J. Pers. Soc. Psychol.* **103**, 54–69. (doi:10.1037/a0028347)
8. Christensen RHB. 2019 ordinal - Regression Models for Ordinal Data. R package version 2019.3–9. https://cran.r-project.org/web/packages/ordinal/.
9. R Core Team. 2018 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
10. Westfall J, Kenny DA, Judd CM. 2014 Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.* **143**, 2020–2045. (doi:10.1037/xge0000014)
11. Bates D, Kliegl R, Vasishth S, Baayen H. 2015 Parsimonious mixed models. *arXiv arXiv:1506.04967*.
12. Irwin JR, McClelland GH. 2001 Misleading heuristics and moderated multiple regression models. *J. Mark. Res.* **38**, 100–109. (doi:10.1509/jmkr.38.1.100.18835)
13. Maxwell SE, Delaney HD. 1993 Bivariate median splits and spurious statistical significance. *Psychol. Bull.* **113**, 181–190. (doi:10.1037/0033-2909.113.1.181)