PSYCHOLOGY

The Dunning-Kruger effect revisited

The Dunning–Kruger effect describes a tendency for incompetent individuals to overestimate their ability. The effect has both seeped into popular imagination and been the subject of scientific critique. Jansen et al. combine computational modelling with a large-scale replication of the original findings to shed new light on the drivers of the Dunning–Kruger effect.

Matan Mazor and Stephen M. Fleming

In one of the most highly replicable findings in social psychology, Kruger and Dunning[1] showed that participants who performed worse in tests of humour, reasoning, and grammar were also more likely to overestimate their performance. In their original report, Kruger and Dunning interpreted this overconfidence in the self-reports of low performers as a metacognitive deficiency, such that poor performers suffer a 'dual burden': in addition to their incompetence in the task, they are unable to identify their own errors. As the title of the original paper put it, poor performers are both "unskilled and unaware of it." A new study in *Nature Human Behaviour* now puts that explanation to the test[2].

The classic metacognitive interpretation of the Dunning–Kruger effect has been challenged by alternative explanations. Krueger and Mueller[3] suggested that the observed overconfidence of poor performers is an instance of regression to the mean: the statistical tendency of extreme samples (here, poor performers) to move toward the group mean when resampled (here, in the form of retrospective performance evaluation). This interpretation identified the origin of the Dunning–Kruger effect in the statistical reasoning of scientists, rather than in the ratings of participants in the task.

In their paper, Jansen and colleagues point out that such regression to the mean can emerge not only as a statistical artefact of data analysis, but also due to the influence of prior beliefs within individual rational observers[2]. To see this, imagine that all participants approach a quiz with a prior belief that they will be correct around 70% of the time. After providing their answers, participants then estimate how well they thought they did. Since no feedback is delivered during the quiz itself, they must rely on noisy confidence signals to evaluate their performance. A high-performing participant who was objectively correct on 7 out of 10 questions may be certain they got 5 questions right, 2 questions wrong, but be unsure about the remaining 3 questions. This participant can go on to accurately estimate that they got 7 out of 10 questions right overall, which is both consistent with their prior belief as well as with their internal 'data' (confidence signals) about their performance. In contrast, a low-performing participant with only 4 correct responses may be confident that they got 2 questions right, 5 questions wrong, and also be unsure about the remaining 3 questions. For this participant, it would be rational to combine these confidence signals with their prior belief to give an estimate of 5 correct responses—an overestimation of their actual performance. In Bayesian reasoning, this is known as 'shrinkage': the rational attraction of surprising samples to the prior mean. This suggests that one explanation of the Dunning–Kruger effect is that it reflects rational Bayesian inference: low performers will appear to overestimate their performance because the noisy data is not enough to override a prior expectation that they will perform well.

To formalise this idea, Jansen and colleagues built a computational model in which rational subjects have access to a noisy internal representation of response accuracy[4]. Simulations of this rational Bayesian model indeed gave rise to a Dunning–Kruger effect at the group level, with pronounced overestimation of task performance in low performers. Notably, here the effect cannot reflect metacognitive incompetence, as the model assumes identical insight across performance levels. An alternative model is that there is a systematic correlation between low performance and metacognitive incompetence. Jansen et al. simulated different versions of this model as well. Predictions of the two model families mostly agreed, but diverged for participants with either very poor or very high performance. By collecting a large sample of online participants, they were able to zoom in on these tails of the performance distribution and establish that the data are more consistent with the second, performance-dependent model. In other words, not only is the Dunning–Kruger effect not merely a statistical artefact at the group level, it also cannot be explained solely by Bayesian shrinkage in the rational estimations of individual participants.

Can we conclude that the Dunning–Kruger effect is metacognitive in nature? The answer to this question depends on what we mean by 'metacognitive'. If we mean that participants with low performance also have a noisier representation of their accuracy, the answer is yes. Jansen and colleagues' model fits make clear that the data are best captured by a performance-dependent change in estimation noise. However, the reasons for this change in estimation noise remain to be determined. Many process models of decision-making, including signal-detection and evidence-accumulation models, naturally predict that low performance should be accompanied by higher levels of uncertainty about task accuracy without postulating additional metacognitive factors[5]. According to this interpretation, the Dunning–Kruger effect is not necessarily the signature of a double burden, but may instead be the signature of a single burden that manifests itself in two ways: in task performance and in performance estimation. Alternatively, changes in estimation accuracy may stem from a second-order process that is distinct from processes driving task performance[6]. Disentangling these alternatives may be possible by collecting data on confidence in individual trials, in addition to global performance estimates, to ask whether metacognitive 'efficiency'—metacognitive noise corrected for task performance—is itself altered in the tails of the distribution[7,8].

In addition to its important theoretical contribution, this study is a beautiful example of the progress made by the field of cognitive science since the turn of the century, when the original Dunning–Kruger paper was published. Careful computational modelling has allowed Jansen and colleagues to identify a diagnostic property of two model families that make different assumptions about latent cognitive variables. Large-scale online data collection has made it possible to generalize the results to populations more diverse than psychology undergraduates and, for the first time, to reliably quantify effects that rely on precise estimates of the tails of a distribution. Finally, preregistration and open data-sharing now make it possible for other researchers to transparently interrogate the results and to easily build on and extend this work. ❒

Matan Mazor[1]✉ and
Stephen M. Fleming[1,2,3]

[1]Wellcome Centre for Human Neuroimaging, University College London, London, UK. [2]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK. [3]Department of Experimental Psychology, University College London, London, UK.

✉*e-mail: mtnmzor@gmail.com*

Published online: 8 April 2021
https://doi.org/10.1038/s41562-021-01101-z

References

1. Kruger, J. & Dunning, D. *J. Pers. Soc. Psychol.* **77**, 1121–1134 (1999).
2. Jansen, R. A., Rafferty, A. N. & Griffiths, T. L. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-021-01057-0 (2021).
3. Krueger, J. & Mueller, R. A. *J. Pers. Soc. Psychol.* **82**, 180–188 (2002).
4. Burson, K. A., Larrick, R. P. & Klayman, J. *J. Pers. Soc. Psychol.* **90**, 60–77 (2006).
5. Vuorre, M., & Metcalfe, J. *Metacog. Learn.* https://doi.org/10.1007/s11409-020-09257-1 (2021).
6. Fleming, S. M. & Daw, N. D. *Psychol. Rev.* **124**, 91–114 (2017).
7. Fleming, S. M. & Lau, H. C. *Front. Hum. Neurosci.* **8**, 443 (2014).
8. McIntosh, R. D., Fowler, E. A., Lyu, T. & Della Sala, S. *J. Exp. Psychol.* **148**, 1882–1897 (2019).

Competing interests

The authors declare no conflicts of interest.