PERSPECTIVE
https://doi.org/10.1038/s41562-021-01203-8

nature human behaviour

Aspiring to greater intellectual humility in science

Rink Hoekstra [1,4] ✉ and Simine Vazire [2,3,4]

The replication crisis in the social, behavioural and life sciences has spurred a reform movement aimed at increasing the credibility of scientific studies. Many of these credibility-enhancing reforms focus, appropriately, on specific research and publication practices. A less often mentioned aspect of credibility is the need for intellectual humility or being transparent about and owning the limitations of our work. Although intellectual humility is presented as a widely accepted scientific norm, we argue that current research practice does not incentivize intellectual humility. We provide a set of recommendations on how to increase intellectual humility in research articles and highlight the central role peer reviewers can play in incentivizing authors to foreground the flaws and uncertainty in their work, thus enabling full and transparent evaluation of the validity of research.

A recent editorial in *Nature Human Behaviour*[1] laments the fact that academia prefers clear and polished stories over honest but less clear-cut ones, with "research projects [presented] as conclusive narratives that leave no room for ambiguity or for conflicting or inconclusive results". Although honesty and clarity are not mutually exclusive, the pressure resulting from this probably affects the quality and validity of our scientific work. Most journals still seem to favour clear stories despite an arguably larger risk of these having validity issues. This presents a real dilemma for authors—especially those who do not have the luxury of a tenured position—who would like to put intellectual humility front and centre. That is, to increase the odds of getting a publication, they are encouraged to present their stories as better than they actually are. Being more honest or humble probably has a negative effect on their CVs. In the following, we present a constructive solution for this apparent stalemate, which reverses the reward structure in such a way that authors are encouraged to write papers that 'tell it like it is'.

Part of a well-functioning and trustworthy science is communicating in a well-calibrated way[2]. When scientific communication is overconfident or contains too many exaggerations, the field stands to lose its credibility, even if the methods and statistics underlying the research are sound. Thus, a core value of the scientific ethos should be intellectual humility, whereby there is a commitment to foregrounding flaws and uncertainty. We define intellectual humility in a similar way to Whitcomb and colleagues[3], who emphasize "owning our limitations". According to this definition, "owning one's intellectual limitations *characteristically* involves dispositions to: (1) believe that one has them; and to believe that their negative outcomes are due to them; (2) to admit or acknowledge them; (3) to care about them and take them seriously; and (4) to feel regret or dismay, but not hostility, about them"[3]. When applied to scientists and their reporting of their scientific work, this suggests an emphasis on owning the limitations of our work by being transparent and non-defensive about them.

According to this definition, transparency is a necessary component of intellectual humility in the context of research outputs because we cannot own the limitations of our work without being transparent about them. But intellectual humility in research papers is about more than transparency because owning our limitations is more than just acknowledging them; it entails a commitment to foregrounding them, taking them seriously and accepting their consequences.

In this and most other conceptualizations and operationalizations (for example, questionnaire measures of intellectual humility), the opposite of intellectual humility is characterized by intellectual arrogance[4,5], overconfidence[6] and intellectual fragility[4,6]. While these definitions and measures were not specifically developed in the context of scientists sharing their work, many of the specific components have clear implications for scientific writing. For example, Alfano and colleagues'[4] measure of intellectual humility includes a facet called 'corrigibility' (for example, 'I appreciate being corrected when I make a mistake'), which seems especially relevant for the context of presenting scientific results. In addition, Van Tongeren and colleagues[7], in discussing intellectual humility as a facet of humility, point out that intellectual humility involves "present[ing] one's ideas in a modest and respectful manner".

This theme of intellectual humility versus arrogance is also present in Merton's[8] description of the ethos of science, particularly his description of the norm of disinterestedness. In describing the temptation to exaggerate, Merton writes "myths will seem more plausible and are certainly more comprehensible to the general public than accredited scientific theories […]. The borrowed authority of science bestows prestige on the unscientific doctrine"[8]. Merton, however, may have been too optimistic to think that this temptation is kept in check by sanctions from scientific peers and the psychological conflict of violating one's professional ethos.

In contrast to the prescriptions of intellectual humility, scientists overclaim — and this problem may even have gotten worse over time. In an analysis of all scientific abstracts published in PubMed between 1974 and 2014, Vinkers and colleagues[9] found that the frequency of positive words substantially increased over time. Although this does not necessarily entail intellectual arrogance, it is probably not an overstatement that scientific humility is practised less than it is preached[10].

Why has this exaggeration gone unchecked? First, we believe that authors and journal editors are often aligned in their motivation to publish bold claims, sometimes outstripping the evidence to support them. In particular, in our opinion, editors may reward rather than penalize authors for overclaiming because editors often lack the information necessary to gauge the validity of the claim. Moreover, they may believe their journal will achieve a higher journal impact factor if they publish articles with bold rather than circumscribed claims, if they assume, as we do, that bolder claims are more likely to capture readers' attention and thus get cited. However, that there

[1]University of Groningen, Groningen, The Netherlands. [2]University of California, Davis, Davis, CA, USA. [3]University of Melbourne, Melbourne, Victoria, Australia. [4]These authors contributed equally: Rink Hoekstra, Simine Vazire. ✉e-mail: R.Hoekstra@rug.nl

is some evidence to suggest that this assumption may not hold[11]. Thus, the incentives for both authors and journal editors may be aligned in favour of bold claims at the expense of accuracy, particularly when claims can be published without all of the information necessary to verify them.

Although there is a pervasive tendency to overclaim in their own manuscripts, the value of humility seems to be widely supported among scientists[12], with only some openly valuing boldness. Mitroff[13], for example, argued that boldness (rather than humility) leads to claims that are more easily testable, and falsifiable, than weak, circumscribed claims. This is indeed a virtue of boldness, but only when these bold claims are presented clearly as speculation, put to empirical tests and actually abandoned when falsified, all of which require a degree of intellectual humility. Only then can a fruitful environment be created for bold, falsifiable claims to be put forward and (usually) refuted.

Another argument against adopting a more humble approach to scientific claims is its potential negative impact on clarity. Arguably, a more nuanced message is harder to get across than an overly simplistic one. However, in science, as opposed to fiction, clarity must be earned. When our understanding of a phenomenon is unclear and incomplete, we have no choice but to admit to those gaps in our understanding. As Einstein is alleged to have said[14], "Everything should be made as simple as possible, but not simpler."

A related argument is that writing an intellectually humble scientific paper can be hard to do within the word limits imposed by many scientific journals. In some cases, it is true that intellectual humility would require extra content, although we suspect that this argument is used more often than is justified. Moreover, we would argue that if authors cannot make limitations and uncertainty prominent while staying within a journal's word limits, then the manuscript is not a good fit for that journal. We hope journals will consider whether they are selecting the highest quality work or encouraging authors to cut corners and submit lower quality work by imposing strict word limits, but ultimately the onus is on authors to submit work that is scientifically responsible.

In short, we think that although strong testable hypotheses and clarity are important for science, intellectual humility is rarely at odds with those values and should be re-established as a key scientific virtue. This is also reflected in how we educate our students: most curricula describe 'the scientific method' as emphasizing self-correction, but the importance of being careful and calibrated—of owning the limitations of our work—is seldom explained. Moreover, it is countered by contradictory training that encourages students to write up their work by telling the most convincing story possible, thus sweeping shortcomings under the rug (see, for example, D. Bem's *How to write an empirical article*[15], which we see as basically reducing scientific writing to a rhetorical game). Such lessons may unfortunately be more the rule than the exception.

Researchers also receive mixed messages regarding intellectual humility from gatekeepers such as journal editors and grant-funding agencies, who often explicitly demand 'groundbreaking' or 'transformative' work. For example, in a 2012 column reflecting on his 5 years as editor-in-chief of *Psychological Science*, R. Kail[16] described the most common reason for rejecting manuscripts without review as follows:

"The Pink Floyd Rejection: most triaged papers were of this type; they reported work that was well done and useful, but not sufficiently groundbreaking. So the findings represented just *another brick in the wall* of science" (emphasis in original).

Truly groundbreaking research, however, should be rare. Thus, editors are probably incentivizing exaggerated claims of how groundbreaking the study is, especially if journals do not require authors to report their work so transparently that the credibility of the groundbreaking claims can be thoroughly assessed. When authors can omit (intentionally or not) relevant information that would be necessary to assess the credibility of the claims, then rewarding groundbreaking claims amounts to rewarding exaggeration and punishing transparency.

In the following sections, we argue that peer reviewers are in a pivotal position to promote intellectual humility with little or nothing to lose. We then describe how intellectual humility can be promoted in research articles.

The power of reviewers

The researcher who wishes to act in accordance with intellectual humility by fully owning the limitations of their work is probably putting herself at a competitive disadvantage compared with others who adhere to this norm less strictly. This can be especially problematic for early-career researchers. Ultimately, the responsibility for prioritizing intellectual humility is shared by many stakeholders, including authors, journals, hiring and promotion committees, funding agencies and science journalists. For example, editors and publication committees who want to incentivize intellectual humility in their journals could implement policies that make it clear to authors and reviewers that owning the limitations of one's research will be considered a prerequisite for publication rather than a possible reason to reject a manuscript. In addition, journals should regularly evaluate whether these policies are actually adopted in practice. This will take time and effort, but will eventually lead to a more robust and credible journal. As new and better metrics are developed to track journal quality (for example, TOP Factor[17]) that are not primarily a measure of popularity or citation impact, we hope that the reputations of journals will be more closely linked to their performance on these kinds of credibility-enhancing measures.

In the meantime, however, we believe that there is, fortunately, a more immediate way to alleviate this apparent stalemate, and it lies in the hands of those who endorse the norm of intellectual humility but whose position or income is independent of enforcing it: reviewers. As reviewers are researchers themselves, they probably endorse this norm (assuming that the findings by Anderson et al.[12] can be generalized), and their job or income is not affected by the reviews they write. Hence, they are in a pivotal position to actually incentivize intellectual humility with little or nothing to lose. Of course, their calls for more humility could be ignored by the handling editor, but in that case the editor would typically have to actively argue against this, which is far more difficult than the common practice of simply ignoring the issue of intellectual humility. And as most researchers also regularly review, this provides all of us with the power to effect change. This idea is not new: in an article[18] on promoting replicability and transparency in psychological science, the authors state that "Reviewers of journal submissions wield considerable power to promote methodological reform"[18], and they urge reviewers to use this power to push for higher standards, including greater intellectual humility. Similarly, the Peer Reviewers' Openness Initiative[19] focuses on the pivotal role of reviewers in promoting openness. Why not use the same power to promote intellectual humility? Of course, the central role of reviewers does not discharge people in other roles (authors, editors and funding agencies) of their responsibility to contribute to improving intellectual humility in the scientific discourse. Here, we are focusing on reviewers because we recognize the pragmatic constraints, and reviewers have less to lose. Our recommendations, as written, can just as easily be adopted by authors and editors as by reviewers. Later, we also discuss the unique role of authors in the post-publication process.

Reviewers can turn the tables: researchers who are asked by a reviewer to 'tell it like it is' suddenly have a clear incentive to adjust their papers. Whereas they might previously have believed that adding hedges lowers their chances of a publication, they now, at a minimum, need to address why they are not appropriately cautious, which will often result in them making adjustments to satisfy the reviewer. Sometimes, these adjustments may even be in

Table 1 | Recommendations for increasing humility in scientific articles

0. Title and abstract	0.1. The abstract should describe the limitations of the study and boundary conditions of the conclusion(s)
	0.2. Titles should not state or imply stronger claims than are justified (for example, causal claims without strong evidence)
1. Introduction	1.1. The novelty of research should not be exaggerated
	1.2. Selective citation should not be used to create a false sense of consistency or conflict in the literature
2. Methods	2.1. The methods section should provide all the details that a reader would need to evaluate the soundness of the methods and to conduct a direct replication
	2.2. The timing of decisions about data collection, transformations, exclusions and analyses should be documented and shared
3. Results	3.1. Detailed information about the data and results (including informative plots and information about uncertainty) should be provided
	3.2. It should be transparent which analyses were planned and where those plans were documented; weaker conclusions should be drawn to the extent that analyses were susceptible to data-dependent decision-making
	3.3. Inferential statistics should not be used in a way that exaggerates the certainty of the findings; alternatives to dichotomous tests should be considered
4. Discussion	4.1. The statistical uncertainty of results should be incorporated into the narrative conclusions drawn from the results
	4.2. The research summary should capture the full range of results (for example, include our 'most damning result')
	4.3. Causal claims should be only as strong as the internal validity of the study allows
	4.4. Claims about generalizability should be only as strong as the sampling of participants, stimuli and settings allows
	4.5. All conclusions should be calibrated to the confidence in the construct validity of the measures and manipulations
	4.6. Alternative interpretations should be presented in their strongest possible form ('steelmanned')
	4.7. A discussion of the limitations should be incorporated throughout the discussion section, rather than bracketed off in a subsection
5. Post publication guidance for authors	5.1. Insist that press releases and reporters capture the limitations of the work, and correct outlets that exaggerate or misrepresent
	5.2. Encourage criticism, correction and replication of the work, and respond non-defensively when errors or contradictory evidence are brought to light
	5.3. When appropriate, retract papers, issue corrections or publish 'loss of confidence' statements

line with what the authors would have liked to say, but were afraid that it would be held against them. We ourselves have seen several such cases in our experience as authors, reviewers and editors. Sometimes, a reviewer raising these issues may be giving cover to one co-author who had lobbied for greater intellectual humility but lost out to their other (often more senior) co-authors. Indeed, as this manuscript was under review, one of us had a reviewer point out to us, in a review of a different manuscript, that we could do more to foreground the limitations of our design (specifically, in the title). Having this advice come from a reviewer was refreshing and liberating. As authors, we had more confidence that the editor was unlikely to penalize us for taking this suggestion on board and presenting our work in a more circumscribed way. In short, when reviewers push for authors to own their limitations, this is not only good for science but may even be in line with (some) authors' ideals. Of course, reviewers can also use positive comments to reshape incentives; that is, when authors go the extra mile to put the limitations of their work front and centre, reviewers can point this out. This will not only serve as a reward to authors but may also act as a buffer against other reviewers who encourage or demand more self-aggrandizement from the authors. Below, we present a preliminary list of steps we think researchers can take when reviewing a manuscript to increase its intellectual humility.

Intellectually humble research articles

Title and abstract. The title and abstract are perhaps the parts of a manuscript where intellectual humility is the most important. If an author exaggerates or makes unsubstantiated claims in the title or abstract, they risks misleading many people as these are probably the most read parts of a paper. However, because the title and abstract are also what many readers will use to decide whether to read the paper in greater depth, authors are motivated to make the title and abstract catchy. One tempting way to catch readers' attention is by making claims that are bolder than the evidence warrants[9]. We may be hesitant to showcase the limitations of our study or to emphasize what we still do not know in the abstract, but that is what intellectual humility would require of us (item 0.1 in Table 1). Similarly, the title should not state or imply stronger claims than are justified by the evidence (for example, implying a causal association when the evidence is weak or suggesting that a medicine or intervention may be effective in humans without mentioning that the study was conducted using animals; item 0.2 in Table 1).

Introduction. Intellectual humility is relevant even before describing a study and its findings. One function of introduction sections is to motivate the research we will be presenting, and it is also easy to exaggerate or mislead here. First, we may be tempted to exaggerate

the novelty of our work by overplaying the relevance (or existence) of the work that has come before. We may create the illusion of a substantial gap in the literature where one does not exist, or exaggerate the differences between our work and the existing literature (item 1.1, Table 1). Second, we may present an overly strong theoretical and empirical case for our hypothesis or methods, thereby giving a false sense of consistency in the literature by failing to cite contradictory articles (item 1.2, Table 1).

Methods. An intellectually humble methods section is one that gives readers all of the ammunition they need to critique the method. We should not ask our readers to trust us to have chosen good samples, materials and procedures. Thus, to keep ourselves honest and accountable, we should provide all of the information a reader would need to decide for themselves whether our methods were sound. Simmons and colleagues' "21 word solution"[20], which suggested adding the sentence "We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study"[20] to every paper is an important step in this direction, but we should also provide any details about the sampling plan, design, materials and procedures that are relevant to the validity of the study (item 2.1, Table 1). Transparent reporting of the methods is also important because it gives other researchers the information they need to conduct independent, direct replications of our work. Encouraging such verification and correction of our work by others is another hallmark of intellectual humility. That is, to show that we truly care about the limitations of our work, we should encourage and welcome critical attention such as attempts to replicate or reproduce it.

One piece of information that is often relevant for evaluating scientific evidence is the timing of decisions about what to measure and analyse. Thus, documenting and sharing the timing of these decisions (for example, the planned sample size and any deviations from that plan and when and how those were decided) is often necessary for readers to have all the information they need to evaluate whether our data support our conclusions. This means being transparent about which decisions related to data collection, transformations, exclusions and analyses were planned, and which decisions were made after having seen the data (item 2.2, Table 1). This is commonly known as a pre-analysis plan or a pre-registration, and when the plans are submitted for peer review before the study is carried out, this is known as a Registered Report (see, for example, ref. [21]). Although Registered Reports are not offered by all journals, the list is steadily growing (a regularly updated list can be found at https://cos.io/rr/).

Results. Humility can also play an important role in the results section. First, it is important to be as transparent and complete as possible, with a full report of descriptive results and insightful graphical displays of the data[22]. The reporting of results can be done in a more or less intellectually humble way. Choosing to give more information (for example, plotting individual data points and emphasizing uncertainties) is the more intellectually humble approach (item 3.1, Table 1). Because we may not be aware of all of the limitations of our work, owning our limitations means making it as easy as possible for others to detect new flaws. Data and information scientists are developing increasingly clear ways of communicating results along with their uncertainty[23]. Moreover, discussing to what extent assumptions seem to have been violated (or not) should be an integral part of every results section. When our results and conclusions might be dependent on arbitrary choices in data cleaning or analysis (which is often the case), a multiverse analysis[24] might enable us to evaluate the robustness or fragility of the outcomes.

Second, decisions about which results to report should be independent of whether the results confirmed the researchers' expectations: all planned analyses should be reported. Moreover, researchers should make it clear which analyses were planned in advance (and where these plans were documented) and which were data-dependent. There is value to both confirmatory and exploratory (in the narrow interpretation of exploring data without having clear hypotheses) work, but presenting exploratory results as if they were confirmatory is misleading (see also ref. [25]). Moreover, we should calibrate our interpretation of the results appropriately given the degree to which knowledge of the data could have influenced our decisions about which analyses to run or how to specify our analysis (item 3.2, Table 1).

Third, inferential statistics should be a means to quantify uncertainty[26]. In practice, however, statistical results are often used to convince people that certainty about the results is justified (for example, see refs. [27,28]). The commonly used null hypothesis significance testing approach can very easily be used to make a dichotomous result (the result is significant or not) the main conclusion. Despite being very common, this method has often been criticized (see, for example, the editorial "Moving to a world beyond '$P<0.05$"[29] in a special issue of *The American Statistician* on this topic). There are at least two alternatives one could use to acknowledge uncertainty (item 3.3, Table 1). First, instead of presenting the significance of the results, the estimation and precision of the size of an effect could be given a more central position, for example, by means of interval estimates (that is, frequentist confidence intervals or Bayesian credible intervals). Making effect sizes prominent also entails interpreting them, which requires thinking, and relating the outcomes to their practical relevance: seemingly large effects can be practically meaningless, and seemingly small effects can potentially be lifesaving. Simply categorizing effect sizes without reference to context is seldom good advice.

An alternative is to use a Bayesian approach to, for example, present the results for three stereotypical people with different a priori beliefs about the phenomenon in question: the sceptic, the agnostic and the believer (see, for example, ref. [30]). For these three imaginary people, we could describe how the data would affect their posterior beliefs (possibly combined with the Bayes factor to indicate the amount of evidence in the data), assuming that they are updating their beliefs in a rational way. If their posterior beliefs converge, the data are apparently pretty convincing for many people, but if not there is still plenty of room for doubt, at least for some. We realize that actually coming up with prior distributions that describe the beliefs of these three people is easier said than done. Nevertheless, we think it is an idea worth exploring. Finally, one could report results from both frequentist and Bayesian procedures simultaneously[31], thereby showing how different analyses can lead to different (or similar) conclusions.

Discussion. After the title and abstract, the discussion section is probably where most intellectual arrogance creeps into scientific articles. One of the biggest responsibilities that we have in a discussion section is to communicate the doubt surrounding our estimates and conclusions. This means not only communicating the statistical uncertainty but also other sources of epistemic uncertainty. These include doubts about the soundness of the design and operationalizations, the generalizability of the results, and the soundness of any causal inferences, among others.

Statistical uncertainty. The strength of the conclusions presented in a manuscript should decrease as the potential for error and bias in our results increases. Often, authors leave the statistical uncertainty behind when moving from the results section to the discussion section, but the precision and bias in estimates continue to be relevant when giving verbal descriptions of our results and their implications. To the extent that we obtained noisy or biased estimates, or that we failed to constrain all potential flexibility in our collection and analysis of data (that is, had less than a thorough and detailed

pre-registration that was exactly followed), our conclusions should often be more circumscribed (item 4.1, Table 1). Furthermore, when summarizing our results in the discussion section, we should be careful not to cherry-pick only the most compelling results, but present the full range of results that do and do not support our conclusions or, at a minimum, the 'most damning results' (see ref. [32] and item 4.2 in Table 1).

Design limitations. No study design is perfect and it is important to acknowledge how our study design falls short of the ideal. For example, when full experimental control is not possible (and even when it is), we should acknowledge threats to internal validity (item 4.3, Table 1). When we cannot randomly sample from the population to which we wish to generalize, or sample from the full range of stimuli or settings to which we wish to generalize, we should acknowledge threats to external validity[9,33] (item 4.4, Table 1). Related to this, claims about implications for application or interventions require direct evidence in the relevant setting. A common temptation in basic research is to make claims about potential implications of the findings for applied contexts or interventions. While clearly marked speculation about the possibility of such implications down the road is appropriate, we should keep in mind that there are many steps required to go from a basic finding to responsible application[34].

Another common limitation is that we are often faced with trade-offs when it comes to the quality of our measures and manipulations. When we have made compromises to the validity of the measures or manipulations used (for example, using measures with low reliability, poor method match or unknown validity), we should acknowledge threats to construct validity[35] (item 4.5, Table 1). It is fairly common for authors to acknowledge these threats but then fail to adjust their conclusions in light of them. Intellectual humility requires more than cursory statements about these limitations; it requires taking them seriously and limiting our conclusions accordingly.

Alternative interpretations. As authors, we have a responsibility to give readers the full range of possible interpretations of our results, even if we present a case for favouring one interpretation over others. We ought to make the strongest possible case for our non-favoured interpretations (item 4.6, Table 1). Presenting alternative interpretations merely to immediately push them aside does not leave the impression that we are genuinely open to the possibility that our favoured interpretation might be wrong. In some cases, we may even wish to remain agnostic about which interpretation is most likely to be valid or propose future tests that would help to discriminate among the possible explanations.

Do not bracket off limitations. It can be tempting—and even feel noble—to put all of the caveats and limitations in one subsection of the discussion (it can also help disarm future critics and reviewers). However, we often then go on, in other parts of the report, to interpret our results and draw conclusions as if these limitations did not exist. While this strategy is indeed more humble than a discussion section without any mention of limitations, it is not fully 'owning' the limitations of our work. According to Whitcomb et al.'s[3] (and our) definition of intellectual humility, such humility requires more than just acknowledging limitations but foregrounding them and fully accepting their consequences. A well-calibrated discussion section should therefore incorporate the limitations throughout (item 4.7, Table 1) and calibrate the strength of the conclusions with the limitations in mind, rather than bracketing them off.

Summary. The suggestions above are by no means exhaustive, but they could function as guidelines that authors can use when writing a paper and to which peer reviewers could refer when evaluating a paper. Ideally, authors who value intellectual humility could implement these suggestions before submitting a paper, and when they do, reviewers should commend them for doing so. However, we realize that some will be hesitant to adopt these out of fear of reducing the chances of the paper being accepted. Once the paper has been published, however, authors are in the best position to encourage responsible interpretations of their work. We now turn to what authors can do after publication to ensure that they continue to own the limitations of their work.

Post publication. As Schubert[36] writes about in his blog post on 'hedge drift', researchers have a responsibility to anticipate that others, including communication offices putting out press releases[37], are likely to repeat their findings without as many caveats, probably in an exaggerated form. We can prevent this by, for example, insisting that the caveats be reported alongside the exciting claims when reviewing press releases or talking to journalists (item 5.1, Table 1). Although this may result in journalists not reporting on our work, journalists may increasingly see these hedges as vital to responsible reporting of science, and perhaps even become suspicious of scientists who do not emphasize the uncertainties. We can also actively correct misrepresentations and exaggerations of our work, for example, by contacting outlets that get it wrong and asking for a correction.

Another way to exercise intellectual humility after publication of our work is to encourage corrections and criticism, including replication attempts, and to respond non-defensively if others do find flaws in our work or present evidence that undermines our claims (item 5.2, Table 1). When warranted, we should also consider retracting papers, publishing corrections or publishing 'loss of confidence' statements (for example, see ref. [38], which reports a project in which authors were asked which claims they have published that they have lost confidence in) (item 5.3, Table 1).

Concluding remarks

In an ideal world, scientists have an obligation to put the limitations and uncertainty in their findings front and centre. Although there are positive developments, such as journals asking for more honest papers[1] and a growing number of journals accepting Registered Reports (in which the decision to publish a paper is made independent of the results), there are still many forces pushing against intellectual humility. Exaggeration is beneficial for the individual (at least in the short run), but it can be detrimental for the group, and eventually for science as a whole, given that credibility is at stake. The downward spiral in this collective action dilemma is not easily reversed, and some people are in a more secure position to make principled but disincentivized choices. Given that scientists have professional obligations and ethics to live up to, intellectual humility should be a factor in the decisions we all make. Researchers should not hide behind a flawed system, the incentive structure or their lower-ranked position (see, for example, ref. [39]). Although they certainly play a role, these factors do not release us from our moral and professional obligations to own the limitations of our work by prominently featuring those limitations and incorporating their consequences into our conclusions.

But many may feel they are not in the position (yet) to 'do the right thing'. That is an understandable position given the non-scientific interests most of us must factor into our decisions. Luckily, reviewing provides an almost cost-free opportunity for all of us to contribute to incentivizing intellectual humility. Even if we are not always able to apply these practices as authors, out of fear of lowering our chances of success, we can flip the script for the authors whose manuscripts we review and make their success at least partly dependent on the amount of humility they show in their papers, thus resulting in more honest and more credible papers. In our view, this is not hypocritical, but trying to act in accordance with our norms within the limits of our capabilities. Of course, these same reviewers would

have to be ready and willing to accept reviewers' suggestions to be more intellectually humble if the roles are reversed. We suspect that most people who are willing to promote intellectual humility as reviewers would be happy to do so when they find themselves on the receiving end of similar reviews.

Putting ideals into practice will not be easy, but bending with the opportunistic winds that have been blowing in science for decades may eventually prove to be even more costly. Eventually, the decision about whether and how to live up to our principles is up to each of us, but we agree with Yarkoni's blog post[39] that we should be able to justify what we do and do not do. Moreover, the costs of sacrificing our principles may be less salient than but just as serious as the more immediate costs of sticking to them. Although personal benefits may be smaller in the short term, eventually it will be in all our interests to not only mention limitations but also to put them front and centre and to encourage others to critique and correct us, thus getting out of this prisoner's dilemma. As preachy as they sound, our recommendations come from personal experience struggling with balancing intellectual humility and professional incentives, and feeling liberated as authors when reviewers and editors reward humility rather than grandstanding. We hope that by pushing each other to own the limitations of our work, we can reshape the incentives so that intellectual humility is rewarded and the credibility of science increases.

Received: 12 June 2020; Accepted: 18 August 2021;
Published online: 28 October 2021

References

1. [Editorial]. Tell it like it is. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-020-0818-9 (2020).
2. Pashler, H. & De Ruiter, J. P. Taking responsibility for our field's reputation. *APS Letter/Observer Forum* https://www.psychologicalscience.org/observer/taking-responsibility-for-our-fields-reputation (2017).
3. Whitcomb, D., Battaly, H., Baehr, J. & Howard-Snyder, D. Intellectual humility: owning our limitations. *Phil. Phenom. Res.* https://doi.org/10.1111/phpr.12228 (2015).
4. Alfano, M. et al. Development and validation of a multi-dimensional measure of intellectual humility. *PLoS ONE* **12**, e0182950 (2017).
5. Leary, M. R. et al. Cognitive and interpersonal features of intellectual humility. *Pers. Soc. Psych. Bull.* **43**, 793–813 (2017).
6. Krumrei-Mancuso, E. J. & Rouse, S. V. The development and validation of the Comprehensive Intellectual Humility Scale. *J. Pers. Assess.* **98**, 209–221 (2015).
7. Van Tongeren, D. R., Davis, D. E., Hook, J. N. & Witvliet, C. V. Humility. *Curr. Dir. Psych. Sci.* **28**, 463–468 (2019).
8. Merton, R. K. Science and technology in a democratic order. *J. Leg. Pol. Soc.* **1**, 115–126 (1942).
9. Vinkers, C. H., Tijdink, J. K. & Otte, W. M. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ* **351**, h6467 (2015).
10. Yarkoni, T. The generalizability crisis. *Behav. Brain Sci.* https://doi.org/10.1017/S0140525X20001685 (2020).
11. Riddle, T. Linguistic overfitting in empirical psychology. Preprint at https://doi.org/10.31234/osf.io/qasde (2018).
12. Anderson, M. S., Martinson, B. C. & De Vries, R. Normative dissonance in science: results from a national survey of US scientists. *J. Empir. Res. Hum. Res. Ethics* **2**, 3–14 (2007).
13. Mitroff, I. I. Norms and counter-norms in a select group of the Apollo moon scientists: a case study of the ambivalence of scientists. *Am. Soc. Rev.* **39**, 579–595 (1974).
14. Sessions, R. How a 'difficult' composer gets that way. *New York Times* (8 January 1950).
15. Bem, D. J. in *The Compleat Academic: A Practical Guide for the Beginning Social Scientist* (eds Darley, J. M. et al.) Ch. 10 (American Psychological Association, 2004).
16. Kail, R. V. Reflections on five years as editor. *APS Observer* https://www.psychologicalscience.org/observer/reflections-on-five-years-as-editor (2012).
17. Nosek, B. A. et al. Promoting an open research culture. *Science* **348**, 1422–1425 (2015).
18. Davis, W. E. et al. Peer-review guidelines promoting replicability and transparency in psychological science. *Adv. Methods Pract. Psychol. Sci.* **1**, 556–573 (2018).
19. Morey, R. D. et al. The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *R. Soc. Open Sci.* **3**, 150547 (2016).
20. Simmons, J. P., Nelson, L. & Simonsohn, U. A 21 word solution. *Soc. Pers. Soc. Psych. Dial.* **26**, 4–7 (2012).
21. Chambers, C. D. Registered Reports: a new publishing initiative at *Cortex*. *Cortex* **49**, 609–610 (2013).
22. Loftus, G. R. A picture is worth a thousand *P* values: on the irrelevance of hypothesis testing in the microcomputer age. *Behav. Res. Meth. Inst. Comp.* **25**, 250–256 (1993).
23. Dragicevic, P., Jansen, Y., Sarma, A., Kay, M. & Chevalier, F. Increasing the transparency of research papers with explorable multiverse analyses. In *Proc. 2019 Conf. on Human Factors in Computing Systems* (eds Brewster, S. et al.) Paper no. 65 (Association for Computing Machinery, 2019).
24. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).
25. Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. & Kievit, R. A. An agenda for purely confirmatory research. *Perspect. Psychol. Sci.* **7**, 632–638 (2012).
26. Hoekstra, R., Finch, S., Kiers, H. A. L. & Johnson, A. Probability as certainty: dichotomous thinking and the misuse of *P* values. *Psychon. Bull. Rev.* **13**, 1033–1037 (2006).
27. Van der Bles, A. M. et al. Communicating uncertainty about facts, numbers and science. *R. Soc. Open Sci.* **6**, 181870 (2019).
28. Hoekstra, R. Risk as an explanatory factor for researchers' inferential interpretations. *Math. Enthus.* **12**, 103–112 (2015).
29. Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. Moving to a world beyond '*P* < 0.05'. *Am. Stat.* **73**, 1–19 (2019).
30. Etz, A. & Vandekerckhove, J. Introduction to Bayesian inference for psychology. *Psychon. Bull. Rev.* **25**, 5–34 (2018).
31. Ly, A. et al. Bayesian reanalyses from summary statistics: a guide for academic consumers. *Adv. Methods Pract. Psychol. Sci.* **1**, 367–374 (2018).
32. Vazire, S. Editorial. *Soc. Psychol. Personal. Sci.* **7**, 3–7 (2016).
33. Simons, D. J., Shoda, Y. & Lindsay, D. S. Constraints on generality (COG): a proposed addition to all empirical papers. *Perspect. Psychol. Sci.* **12**, 1123–1128 (2017).
34. IJzerman, H. et al. Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* **4**, 1092–1094 (2020).
35. Flake, J. K., & Fried, E. I. Measurement schmeasurement: questionable measurement practices and how to avoid them. Preprint at https://doi.org/10.31234/osf.io/hs7wm (2019).
36. Schubert, S. Hedge-drift and advanced motte-and-bailey. *LessWrong Blog* https://www.lesswrong.com/posts/oMYeJrQmCeoY5sEzg/hedge-drift-and-advanced-motte-and-bailey (2016).
37. Sumner, P. et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ* **349**, g7015 (2014).
38. Rohrer, J. M. et al. Putting the self in self-correction: findings from the loss-of-confidence project. *Perspect. Psychol. Sci.*, https://doi.org/10.1177/1745691620964106 (2021).
39. Yarkoni, T. (2018). No, it's not the incentives—it's you. *Talyarkoni Blog* https://www.talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/ (2018).

Acknowledgements

We thank A. Allard, A. Holcombe, H. Kiers, L. King, S. Lindsay and D. Trafimow for valuable input for this paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Rink Hoekstra.