Methodology and Research Practice

A Call for Greater Modesty in Psychology and Cognitive Neuroscience

Richard Ramsey[1] [a]

[1] Department of Psychology, Macquarie University, Sydney, Australia

Collabra: Psychology

Exaggerated claims and low levels of reproducibility are commonplace in psychology and cognitive neuroscience, due to an incentive structure that demands "newsworthy" results. My overall argument here is that in addition to methodological reform, greater modesty is required across all levels - from individual researchers to the systems that govern science (e.g., editors, reviewers, grant panels, hiring committees) - to redirect expectations regarding what psychological and brain science can effectively deliver. Empirical work and the reform agenda should pivot away from making big claims on narrow evidence bases or single tools and focus on the limitations of our individual efforts, as well as how we can work together to build ways of thinking that enable integration and synthesis across multiple modalities and levels of description. I outline why modesty matters for science including the reform agenda, provide some practical steps that we can take to embrace modesty, rebut common misconceptions of what modesty means for science, and present some limitations of the approach. Ultimately, by presenting a more sober view of our capacities and achievements, whilst placing work within a wider context that respects the complexity of the human brain, we will bolster the fidelity of scientific inference and thus help in a small way to generate a firmer footing upon which to build a cumulative science.

1. Introduction

A common strategy in business and journalism is captured by the phrase "simplify, then exaggerate" (Davis, 2017; Herrero, 2014). Most news stories, for example, entail a complex set of interacting pieces that have played out over a relatively lengthy period of history and continue to do so. The journalist's skill is to reduce complex information into something simpler and more digestible, before exaggerating it into a "newsworthy" story that can sell papers and drive clicks online. For a related example, consider common practice in science journalism. Complex health science is often boiled down to whether or not one may die from eating a sausage, thus bringing about the so-called "sausage wars" (Spiegelhalter, 2019b). Such stories guarantee eye-catching headlines by ignoring considerable uncertainty and nuance in order to issue artificially clear-cut health advice (Spiegelhalter, 2019a).

One may think that academia and professional science would not suffer from the same problems because it is unburdened by market forces and the pursuit of sales. A romantic view holds that science is built on different values,

such as integrity and honesty, as well as different systems of operation that mandate a dispassionate, calculated and systematic pursuit of the "truth". However, such a view of science is naïve. The incentive structure of modern science is such that a "simplify, then exaggerate" strategy has become dominant, even if only tacitly. To get published in leading journals, to be awarded grants and to be hired as a postdoc or faculty member, a system-wide bias for novelty, exaggeration and storytelling has emerged (Huber et al., 2019; Nosek et al., 2012). The prizing of novelty over quality represents one overarching driver in the construction of a research culture beset by the widespread use of questionable research practices and low levels of reproducibility (Chambers, 2017; Munafò et al., 2017; Nelson et al., 2018; Open Science Collaboration, 2015; Simmons et al., 2011). Indeed, although there have arguably been recent successes (Shiffrin et al., 2018), many aspects of modern psychology and brain science resemble a creative writing class as much as a systematic science of brain or mind.[1]

Of course, simplification is essential to all science. For a deeper and more complete understanding of any subject, it is common to deconstruct it into pieces, understand each

---

a richard.ramsey@mq.edu.au

[1] Some readers may consider this an unfair characterisation of research in psychology and cognitive neuroscience. I do not for two reasons. With reproducibility levels so low (between 25% and 50% in leading psychology journals, for example: Open Science Collaboration, 2015) and the link between data and inference so poorly aligned in many cases (Yarkoni, 2020), I think this description is justified. Two qualifi-

piece in turn including all interrelations between pieces, and then re-construct it into a complete whole (Gleick, 2011). For example, Isaac Newton made use of simplification when attempting to understand planetary motion by only focussing on the Earth and the Sun while ignoring other known planets of the time (Gleick, 2004). Similarly, in psychological and behavioural science, all computational and statistical models attempt to be useful simplifications of the reality they seek to understand (McElreath, 2020; Smaldino, 2017). Therefore, as I unpack below, I am suggesting that we move towards simplification *without* exaggeration. We should seek to develop a state of science where the fundamental limitations of the general approach are kept front and centre because they are central to the construction of any cumulative science. As such, I am encouraging greater humility, as much as modesty,[2] whereby we strive to place the limitations of our individual efforts within a broader scientific context.

Without the grounding influence that acknowledging fundamental limitations provides, a research field can get carried away on a wave of self-promotion and self-delusion that takes it down a path that is disconnected from the broader scientific endeavour. Although exaggeration is not a new feature of science (Babbage, 1830), the main mission for many psychological and brain scientists increasingly appears to be the demonstration of improbable or even impossible levels of insight and novelty based upon limited means. I do not blame individual researchers. The system has raised researchers on a diet rich with exaggeration, which has normalised a skewed set of expectations and research practices. Indeed, a key tenet of any good science – that inferences should be proportional to the quality of evidence – has become marginalised in a sea of hype. Moreover, the ubiquitous and pervasive nature of hype makes it hard for researchers to even begin to recognise the inherent disconnection between inference and evidence, never mind take action to remedy it. In response, improvements to common research practices have been put forward that aim to place a renewed focus on quality over novelty and exaggeration (Nosek et al., 2012). Like others recently (Hoekstra & Vazire, 2020), my overall argument is that in addition to methodological reform, a spoonful (or bucketful) of modesty is required at all levels and for all involved to help rebalance and redirect psychological and brain science. The unashamed and continual recognition of fundamental limitations – in terms of limits to human mental capacity, as well as limits to modern scientific systems – will, I argue, reduce the temptation for overstatements and thus provide some protection against a weaker and less reliable form of science.

The main body of the current article is organised into four parts. First, I outline why modesty (or lack thereof) matters for science. I provide examples of where immodesty in the form of grand claims on narrow data, which are detached from a wider scientific context, has taken us, and why it presents a substantial and pervasive barrier to progress. I also outline why much of the reform agenda to the reproducibility crisis in psychology suffers from some of the same problems; namely, a plethora of narrow solutions are put forward that make outsized claims regarding a solution to a complex problem. Second, I put forward a few proposals for change that are centred around downplaying estimates of our own abilities and elevating the estimated difficulty of psychological and brain science. I invite greater humility in light of the limited nature of our reasoning abilities (Hintzman, 1991; Kahneman, 2011) and greater respect for the difficulty of the task at hand. Third, I rebut common misconceptions that are often levelled at calls for increased modesty. Such misconceptions typically equate calls for modesty with a call for "boring" science, a lack of ambition, or as a defence for poor research practices. In the fourth part, I raise serious dangers of the approach, which are that it could be costly to career progress or futile in the face of prevailing incentives, whilst also outlining ways to combat and minimise any such negative consequences. Overall, I aim to demonstrate the value of modesty for progress in psychological and brain science, and thereby reinforce the point made previously by researchers and philosophers that intellectual humility and the acknowledgment of limitations is a strong, rather than a weak, scientific stance to adopt (Firestein, 2012; Lilienfeld, 2017; Lilienfeld et al., 2017; Roberts & Wood, 2003; Whitcomb et al., 2015).

2. Why does modesty matter for scientific progress?

2.1 Scientific reform via silver bullets and magic wands

The credibility of psychological science has been questioned recently, due to low levels of reproducibility and the routine use of inadequate research practices (Chambers, 2017; Open Science Collaboration, 2015; Simmons et al., 2011). In response, a "credibility revolution" has begun (Vazire, 2018), which entails wide-ranging reform to scientific practice (e.g., Munafò et al., 2017). The use of questionable research practices emerged hand-in-hand with an almost exclusive focus on novelty in the formal literature, which rewarded impressive-sounding claims based on small amounts of evidence (Nosek et al., 2012). Put another way, relatively general claims are frequently made on narrow evidence, which creates a mismatch between data and inference and reduces the validity of the claims being made

cations are worth emphasising, however. First, I am not referring to all aspects of psychology and brain science research, just a substantial portion. Second, I am not arguing that no progress has been made, just that there is room for major improvement.

2 Although the terms modesty and humility are often used interchangeably, researchers disagree on whether they refer to the same phenomenon or not (Bommarito, 2018). Some suggest that modesty refers to restraint in appearance or behaviour, whereas humility derives from a more grounded perspective, which emphasises one's limitations (Burton, 2018). For the purposes of this paper, ideas surrounding notions of modesty and humility as defined above are equally relevant and therefore I use both terms.

(Yarkoni, 2020; but for a different view, see Lakens, 2020). As Yarkoni (2020) makes clear, sweeping statements about the presumed generalisability of effects are frequently made based upon a small number of studies that used a narrow set of conditions (in terms of participant demographics, stimuli and task manipulations), and which do not mention boundary conditions or provide empirical evidence for the generalisability of the effects. Indeed, there appears to be a misreading of the likely reach and certainty that any one piece of psychological or brain science evidence could possibly provide. The result is a default to exaggeration and a weakening of the link between data and inferences, irrespective of the merits of the chosen methodology. But this state of affairs begs a question: If exaggeration seems widespread and pervasive and may resemble a permanent fixture of modern psychological and brain science, does the reform agenda suffer from the same problem?

To put my cards on the table, I am greatly encouraged by the reform agenda and fully support the general direction of travel (Ramsey, 2020), whilst acknowledging that there is much debate and nuance regarding the merits of the various ways to move forward (Nosek et al., 2019; Rubin, 2017, 2019; Szollosi et al., 2019). Many give due diligence to the multivariate nature of the problem and present a diverse set of solutions (e.g., Munafò et al., 2017; Nosek et al., 2012). Others outline how questionable research practices, such as Hypothesising After the Results are Known (HARKing), come in many forms that vary in how detrimental they are to science (Rubin, 2017). However, it is also common to give the impression that a single bit of reform, which typically has a narrow focus and reflects a solitary tool, resembles the saviour of psychological and brain science. Indeed, many appear to be under the illusion that they have struck gold by uncovering a silver bullet or magic wand. Therefore, there is a sense with which the potency of any single "cure" for science's ills has been exaggerated, much in the same way that empirical findings are frequently overstated.

I contrast this position with a different view. In his outstanding statistics textbook, Richard McElreath emphasises that statistical models are just one tool in a researcher's toolkit and an imperfect tool at that (McElreath, 2020). One of the standout strengths of McElreath's book is that he emphasises that statistical models, however advanced and complex, are fundamentally limited and need to always be framed and used within the wider scientific context. The broader scientific context may entail the importance of theory, the availability of open data and materials, pre-registration, meta-analytical approaches, computational modelling, experimental design, data science and visualisation and many more considerations besides. In stark contrast to recent calls for <insert favourite new reform approach here>, it is the collective that matters most, but often gets ignored. There is a much greater need to develop approaches that synthesise information across different levels of description (e.g., Morton, 2004), as well as provide a systematic structure for psychological research (e.g., Haig, 2014), rather than create hyperbole and unrealistic expectation over one specific tool or endlessly debate a subcomponent of one particular aspect of science, such as p-values (Benjamin et al., 2018; Lakens et al., 2018). Such activities detract from the bigger picture, while also sending the message, whether intended or not, that if only we use p-values appropriately or correct for bias in meta-analyses (for example), all our problems will disappear.

By analogy to medicine, it appears that a single pill is provided, which resembles a magic cure. Such a view, however, does not fit with the multidimensional and multivariate setting that is the "illness" from which our field suffers. The mismatch between problem and solution feels as disconnected as expecting a pill to "cure" Autism Spectrum Condition, which is a heterogenous developmental condition that is likely to result from a complex and multi-factorial underlying causal structure that varies across individuals (Frith, 2003; Happé et al., 2006; Morton, 2004; Plomin, 2018). Or, as an alternative example, it is reminiscent of the ever-growing list of purported quick fixes for effective weight management that ignore the complex biopsychosocial mechanisms that control weight fluctuation (Lean et al., 2018; Raynor & Champagne, 2016). This is not to deny that one tool (e.g., counting calories) could help in some small way to control weight. It also does not deny that it might take a considerable amount of time and effort to make that tool work effectively and understand any associated mechanisms. However, it remains just one tool and one small contribution. One would hope that the presence and visibility of these examples would serve to underscore the inherent difficulty of the task facing psychology and brain science. Instead, however, it seems to galvanise the reporting of success stories. To place undue focus and hope on any one method is to mischaracterise the nature of the problem itself. The problem stems from the complexity of the target we wish to understand, which is the multi-levelled structure of the human nervous system coupled together with the multifaceted nature of human-environment interactions. Therefore, the default assumption and focus should be that a complex problem requires a complex solution.

What follows from this re-positioning of expectations is a pressure to take a step back and try to see how one piece of work may fit into the bigger picture that we may want to understand. Indeed, although much of the reform agenda has focussed on revision to statistical practices (e.g., Cumming, 2012; Gigerenzer, 2018; Lakens et al., 2018), of particular relevance here are recent calls by theoreticians and computational modellers for major improvements to building and specifying theories in psychology (Borsboom et al., 2021; Fried, 2020; Gervais, 2021; Guest & Martin, 2021; Haig, 2014; Muthukrishna & Henrich, 2019; Navarro, 2019, 2021; Oberauer & Lewandowsky, 2019; Robinaugh et al., 2021; Smaldino, 2017, 2020; van Rooij & Baggio, 2021). The logic is that to build a well-made house, you need to work from a clearly specified blueprint (Gray, 2017). In other words, a collection of bricks is no use without a system for putting them into a coherent structure.

One example approach to theory development involves building computational models. Computational models require that relationships between parts of a system are explicitly specified, thus avoiding a sole reliance on narrative descriptions of theories, which are less precise and are harder to interpret, test and falsify (Hintzman, 1991; Reichle, 2020; Smaldino, 2017, 2020). Likewise, firmer claims may be licensed if researchers follow a more systematic approach to the scientific method that spans the whole the-

ory-data cycle, as it can help guide predictions and interpretations (Bassett & Gazzaniga, 2011; Borsboom et al., 2021; Guest & Martin, 2021; Haig, 2014; Robinaugh et al., 2021). For example, several researchers have outlined step-by-step methodologies for transitioning between verbal theories, formal models and the evaluation of data in systematic and principled ways that involve iterative cycles (Borsboom et al., 2021; Guest & Martin, 2021). These frameworks add value by providing researchers with the tools to structure the process of theory development and make more explicit links between theory and data.

The benefits of adopting more formal and systematic scientific processes apply broadly and include how we may think about newly proposed approaches to scientific reform. If one does not have an overarching sense of the scale and sheer number of parts and processes across different levels of a scientific endeavour (some known and some unknown at any point in time), it becomes too easy to exaggerate the value, likely success and overall contribution of any one new tool or approach. Instead, we should appreciate the 'many to many' relationships that exist between scientific tools and outcomes, as well as their interactions. And thus, we should submit to the reality that the brain is not easy to understand, and acknowledge that it might not be built with the requisite structure to understand itself on every level that we wish to understand it on (Bassett & Gazzaniga, 2011; Gazzaniga, 2010; McGinn, 1989).

2.2 The quest for better data and the neglect of value judgments

There is no question that taking steps to gather better data is an important pursuit, which forms the backbone of the credibility revolution (Munafò et al., 2017; Vazire, 2018). What deserves greater recognition, in my view, is that there are limits to what data alone can provide. In this regard, we can learn from economic theory. Economists distinguish between two key factors when arriving at a decision: the quality of the data and an assessment of value (Oster, 2013). The first step involves gathering the right kind of data. This step is typically not straightforward and there is rightly much discussion regarding the merits of different approaches to data collection and analysis in any science. The second step involves evaluating the data based on a wider set of factors such as one's values, priorities and situational conditions. As such, the same data can be judged differently by different people and across different situations. For example, national health policy guidelines for England and Wales are produced by considering the quality of data regarding a particular treatment together with value judgments concerning cost effectiveness and wider societal benefit (Rawlins & Culyer, 2004). Economic theory, therefore, makes clear that cost-benefit trade-offs are an essential part of arriving at an evidence-based judgment.

Considered in this light, value judgments, I would argue, hold important lessons for psychological and brain science, which have been underappreciated to date. To arrive at a judgment about the value of work in psychology or neuroscience, it is not enough to focus only on the kind and quality of data. It is also important to consider a host of wider factors, such as aims and context. For example, one may consider whether the work attempts to understand basic systems or provide more immediate practical benefit, the extent to which the method is particularly novel, labour intensive (e.g., longitudinal designs) or expensive (e.g., fMRI), whether the population is unusual, understudied, difficult to reach or of particular interest (e.g., patients, remote tribes), and many more reasons besides. Each of these considerations could add extra value and be used to justify the research in addition to other concerns about rigour and methodological quality. In some cases, it may be judged that the added value is worth the additional effort or worth compromising on gold standard conventions, such as using a less powerful design.

To illustrate the role that could be played by value judgments I provide a few concrete examples. In the first two cases, value is not clear-cut and instead a judgement needs to be articulated and defended regarding a trade-off between methodological rigour and other dimensions of value. The first example considers the practical constraints of testing an unusual population, such as a patient group. In such cases, a relatively small number of patients may be worth studying because of what they may be able to suggest about the human brain or the patient's condition, which other approaches cannot. But, due to the small sample size and necessary reduction in statistical power or precision, the inferences that are drawn need to be modest, sober and sensible, as well as respect elevated levels of uncertainty. In other words, the conclusions need to be appropriately calibrated.

A second example concerns expensive and/or time-intensive approaches. Cognitive neuroscience researchers who complete intensive training studies across days, weeks and months, for example, require vast hours of training per participant plus repeat and costly neuroscience techniques, such as fMRI (e.g., Apšvalka et al., 2018; Cross et al., 2006). In principle, of course, there is no reason to accept reduced certainty; one would just need to maintain a sufficient level of rigour no matter the type of study. However, in practice, and given finite resources, researchers may need to use a less than optimal design, which sacrifices levels of certainty in the conclusions made. Nonetheless, the researchers may feel that this work can still provide considerable value because it brings insight into neural plasticity in a way in which few other approaches can provide.

In both of these examples, and without the luxury of considerably more resources being available, conclusions are likely to be suggestive rather than convincing and this should be clearly stated. And moreover, it is important to recognise that other researchers may not arrive at the same value judgment. Others may feel that the overall cost-benefit for the advance to knowledge is not satisfying and as a consequence the work is not sufficiently worthwhile. But this is a discussion that is worth having in my view because it is not clear to me that notions of "right" and "wrong" are relevant when it comes to such value judgments. I would argue instead that value judgments should be explicitly stated, justified and then rigorously debated.

A third example of a cost-benefit trade-off worth consideration is the balance between basic research and research with more immediate practical benefit. In many cases, basic research may be worth the effort for the potential advance

to knowledge that it can provide. It is also worth remembering that we cannot know in advance how knowledge gathered in one area of science will ultimately be used by another area of science or how it may one day have real-world impact. In other words, given that it seems misplaced to decide in advance which forms of basic research would lead to practical impact in the long run, I am happy to encourage basic research for the sake of understanding and let time do the rest. However, not everyone has to agree with my value judgments and many have underscored the value of psychological research that provides more immediate practical benefits (e.g., Berkman & Wilson, 2021). What is abundantly clear from this analysis, however, is that a clear justification of value along with a corresponding timeframe needs to be provided as part of the context for interpreting the value of proposed research, as well as research outcomes. There is ample opportunity to provide a broadly-scoped justification of value at stage-1 of the registered report submission process (Chambers, 2013; Nosek & Lakens, 2014), but it would be beneficial if it became routine practice across all aspects of science.

The idea that a wide range of factors need to be considered when assessing the utility and value of research is not at all new (Field et al., 2004; Lakens et al., 2018; Neyman & Pearson, 1933), but it is often overlooked and rarely stated explicitly in empirical papers, which I believe hampers progress. Indeed, at present, there appears to be a default to a more ritualistic approach that is based purely on statistical considerations (Gigerenzer, 2018), whereby the demonstration of high statistical power is enough to demonstrate high value. However, a productive and resource-efficient science cannot only involve confirmatory, pre-registered and high-power studies; it equally needs a range of exploratory and non-confirmatory research approaches (Scheel et al., 2020; Tong, 2019). Indeed, good science is likely to involve a combination of different styles of research, some of which generate hypotheses and some which confirm hypotheses (Nosek et al., 2018). Psychology and cognitive neuroscience researchers may find it easier to avoid hype and overselling if the value of exploratory research, which necessarily leads to weaker and more suggestive inferences, was acknowledged and given the respect that it deserves more broadly.

To be clear, an argument over the relative value of an approach across a range of possible dimensions is not the same as arguing that the data from all approaches are equally valid, effective, complete or robust. I am not, in any way, inviting a lowering of the bar that we set to evaluate the quality of data needed for confirmatory research. I am suggesting that we should spend more time distinguishing between data quality and value judgments. Specific and technical critique remains as valuable and as important as ever. In short, a flawed method is still a flawed method. A method that does not allow even a suggestive claim to be made due to some inadequacy of technical execution or theoretical position still needs to be pointed out and addressed, so that everyone can improve.

I am, however, arguing that it would be a mistake if the pursuit of more rigorous standards had the knock-on consequence of stifling more exploratory, creative, innovative or risky work, as researchers prioritise safe bets and the chance

of a career in science (Lilienfeld, 2017). My view is that we should encourage plurality and variation to develop a broad base of evidence in science, whilst having some clear guidelines on quality. In other words, a strong scientific foundation involves a combination of more exploratory and more confirmatory research (Nosek et al., 2012; Tong, 2019; Wagenmakers et al., 2012). In addition, for some questions, methods or approaches, there may well emerge a fairly universal understanding and acceptance of best practice (e.g., randomised control trials in medical research). At which point, these guidelines should be communicated broadly. I am not arguing against this suggestion in any way. Instead, however, I think there must always be aspects of what are the "right" questions and approaches to study, which cannot be universally agreed upon. We should expect this. We should also expect that different researchers will make different trade-offs between a variety of competing factors when conducting research. These trade-offs, however, should be explicitly and consistently justified. Making these justifications explicit would generate a much stronger science because it would encourage a balance between risk-taking and rigour that ultimately leads to more appropriately calibrated conclusions.

2.3 Complexity demands intellectual humility

Even after data quality and value have been considered together, there remains an even more general impetus for modesty. There is a real possibility that the human brain may be unsuited to understand itself in the ways that we wish to understand it (Bassett & Gazzaniga, 2011; Gazzaniga, 2010; McGinn, 1989). If we consider the well-documented limits and biases in human reasoning (Hintzman, 1991; Kahneman, 2011), plus the complex nature of brain function, as well as the multi-faceted machinery of modern-day science, human cognition simply cannot operate in a bias free manner across the required complexity of information. Set within this context, arriving at "balanced" decisions that consider more than a few factors becomes inherently difficult. Humans make use of mental shortcuts and use heuristics, but the field of interest is broad and diverse, and issues are complicated. And these facts are just not happy bedfellows for ambitious individuals with demanding schedules, so it is not a trivial problem to solve in my view, which is why the role of biases warrants further recognition and consideration (Munafò et al., 2017).

It is also worth remembering that at present, we are only scratching the surface of understanding the true complexity of the brain's functional properties (Cobb, 2020), as well as how such a complexity underpins mental illness (Fried & Robinaugh, 2020). Despite progress decoding mental states via complex machine-learning algorithms (e.g., Huth et al., 2016), the laws that govern brain function may not be discoverable in a human-readable manner anytime soon, if ever. Therefore, common claims that researchers wish to make in psychology and cognitive neuroscience might not be in tune with the methods currently available (Yarkoni, 2020).

Adopting approaches from older sciences, such as physics, appears eminently sensible on first glance. But the structure of older sciences may only offer partial guidance

on how to proceed because of the inherent ramping up of complexity that the human brain presents, which includes variability across individuals and settings, and means there is likely to be a looser reliance on law-like functions (Sanbonmatsu & Johnston, 2019). One consequence might be that decisive physics-like experiments may be impossible to emulate in psychology and we should revise expectations accordingly (Debrouwere, 2020). We may instead need a different vision, which incorporates a different and more modest set of expectations, for how to structure this type of science.

In summary, I am suggesting that to evaluate scientific outputs in psychology and cognitive neuroscience, we need more than good quality data; we must also link data to a more general value system that reflects a cost-benefit analysis in relation to a wider set of factors. The more general point I wish to make here is that whatever one's views regarding the quality of data or the value of a particular approach, it should result in modesty and respectful uncertainty in stating achievements rather than swagger and certainty. We should emphasise that good science reduces uncertainty, but it does not eradicate uncertainty (Spiegelhalter, 2019a). Indeed, any science should expect a state of partial ignorance, rather than downplay it as unusual or unfortunate (Firestein, 2012). The appeal for modesty applies at all levels and to all actors: individuals, research groups, sub-disciplines and the whole field of psychological and brain science. As such, the current proposal resembles a call for "massive modesty" – modesty across every conceivable level and approach.

3. How do we embrace intellectual humility?

Here I outline some general ways to embrace intellectual humility, which can operate across different aspects of the research process from writing papers or grants to developing skills, forming collaborations and designing research programmes. Relatedly, others have recently made similar suggestions, but in a narrower context that is specifically tied to writing and reviewing journal articles (Hoekstra & Vazire, 2020). As such, Hoekstra and Vazire's (2020) proposals are a welcome complement to the proposals below, by providing more detailed and concrete suggestions for one central part of the research process. Yet others have constructed checklists for spotting hype (Meichenbaum & Lilienfeld, 2018), as well as made proposals for how the principle of intellectual humility could anchor entire graduate training programmes in clinical psychology (Lilienfeld et al., 2017). More specifically, Lilienfeld and colleagues (2017) argue that embedding intellectual humility throughout graduate training programmes recognises that we are all susceptible to biases in reasoning and that science can offer some welcome inoculation against them. Such proposals make clear that the implications of embracing intellectual humility run the gamut of science and extend beyond the proposals I make here. Therefore, the ideas below should be considered as entry points to stimulate further discussion, rather than an exhaustive account of ways to embrace intellectual humility.

3.1. Be explicit about aims, value judgments and the generality of claims.

It is an old argument, which many trainee research students will be familiar with, to clearly state aims and make careful and sensible inferences that link together aims, methods and results. Simply put, inferences should be directly proportional to the quality of the evidence. However, this basic tenet of the scientific process needs emphasising and re-stating, in my view, because it has important consequences for setting expectations appropriately and for building a robust and cumulative science. In many cases, a toning down of claims to reflect the narrow nature of the evidence seems appropriate (Yarkoni, 2020). It would also help to guide interpretation, as well as replication and extension efforts, if it became commonplace to explicitly provide constraints on generality, rather than let others guess how far-reaching one may expect the results to be (Simons et al., 2017). As such, there should be a diversity of claims, rather than a monoculture whereby all discussions are stated in equally strong terms, irrespective of the strength and type of data obtained. There should also be a diversity of limitations presented alongside such claims, rather than minimising limitations in the hope that reviewers miss them. Again, all of this may sound familiar, but it is nonetheless a neglected and overshadowed aspect of our profession. Finally, we should place aims and results within a clear value system that provides a wider context and helps to keep claims grounded. For example, it would be useful to see a clear justification for how the work should be considered: to what extent is it more exploratory or confirmatory in nature? Is it probing understanding of basic systems or providing more practical benefit? And over what kind of timeline may such outcomes be realised?

3.2. Upskilling and team science.

One consequence of the credibility revolution is that we need to concurrently learn new skills and build a culture of team science. Given the apparently endless number of new skills that we could learn, it is impractical for everyone to focus on the same skills to the same degree. As such, the complement of upskilling is a team science approach, which places more emphasis on collaboration and building formal teams that have diverse skills. The team science approach reflects a position that shows that we are fundamentally limited and working together may be the only way to scale-up successfully. Team science in this context, therefore, is not meant to represent a buzzword used for grant writing. It represents a fundamental shift in how psychological and brain science research gets done. Good examples would include the reproducibility project (Open Science Collaboration, 2015), the Many Labs projects (e.g., Klein et al., 2014), the psychological science accelerator (Moshontz et al., 2018), as well as adversarial collaborations (Ellemers et al., 2020; Kahneman, 2003) and interdisciplinary collaborations in mental health research (Fried & Robinaugh, 2020).

With limited time and resources, choices need to be made in order to strike a balance between upskilling and building more diverse teams. To provide a concrete example, consider new developments in statistical analysis and

modelling (e.g., Barr et al., 2013; Cumming, 2012; McElreath, 2020). For me, it became essential to delve into a more complex statistical modelling space, which involves Bayesian and multi-level approaches (Kruschke & Liddell, 2018; McElreath, 2020). I remain a novice in this domain, however. Others may make a different judgment. Others may prioritise theory development or computational modelling over more complex statistical modelling and instead choose to setup formal collaborations with statisticians. Either way, whether one upskills or develops a more diverse team, the result is that business as usual is not sufficient. The ritualistic and somewhat mindless use of t-test and ANOVA analyses that adopt a sole focus on *p*-values and ignore multi-level structure must give way to more considered statistical approaches (Gigerenzer, 2018). This much has been known for decades but the tools are now freely available to do so, although they require much more involvement and understanding from the user (Bates et al., 2015; Bürkner, 2017; Kruschke, 2015; McElreath, 2020). As such, hard choices and decisions must be made that involve sustained and continued effort to learn new skills and/or develop new collaborations and more diverse teams that include statisticians, programmers, data scientists, computational modellers and many more working together.

3.3. Respect the bigger picture.

We should expend much more energy building systems that enable ways to synthesise information. Such synthesis can be across multiple levels of description, such as those put forward in developmental science (e.g., Morton, 2004), or multiple studies, such as large-scale aggregation of information across neuroimaging datasets (e.g., Neurosynth.org, the ENIGMA Consortium: http://enigma.ini.usc.edu/, and many others). In addition, rather than take a bespoke or ad hoc approach to the scientific process, which is left implicit and unjustified in most cases, we may want to embrace proposals for a more systematic approach to the theory and method cycle in psychology (e.g., Borsboom et al., 2021; Guest & Martin, 2021; Haig, 2014). Engaging directly with such approaches will help to enforce modesty in our expectations of what individual pieces of work can provide, as well as what any one tool/approach can possibly provide.

The same logic can be applied to the reform agenda also. For example, it may be helpful to explicitly situate a single proposed piece of reform – say, a new statistical approach to hypothesis testing – within a broader context. To do so, one may supplement written arguments, which may be more easily glossed over, with figures, diagrams and formal models of the overall scientific process. The reasoning is the same as that supporting formal modelling approaches in psychology – it makes the account explicit and aids transparency (e.g., Guest & Martin, 2021; Hintzman, 1991; Reichle, 2020; Smaldino, 2017). So, instead of just writing a paper about *p*-values, you also build a model of the scientific process and then situate *p*-values within it. Then it becomes visually and/or computationally clear that the model of science being followed involves 10, 20, 50+ interacting parts and *p*-values are but one part of the statistics subcomponent. It could even be a box and arrow diagram of sci-

ence, just to reinforce the position that many factors are involved. Of course, computational models already exist for how modern science operates (e.g., Smaldino & McElreath, 2016), and these may offer a useful contact point. In short, I think formal modelling approaches that explicitly recognise the wider context and inter-relations between other factors would help avoid situations where reform proposals are superficially characterised in a narrow way by authors or by other scientists.

Moreover, greater recognition of the systems that guide the scientific process in psychological and brain science may emphasise that the structure of an 'old science', such as physics, may not be a completely suitable platform for 'new science'. That is, based on levels of complexity in human systems and behaviour, there is good reason to suggest that we may need to hold different expectations regarding the level of granularity, clarity and definitiveness of our work compared to work on physical systems (Debrouwere, 2020; Sanbonmatsu & Johnston, 2019).

3.4. Re-boot expectations in 'slow science' mode.

Authors, reviewers, editors, grant panels and hiring committees alike need to be far more sensitive to the fact that 'less is more' in science and expectations need to change (Frith, 2020; Krakauer, 2019). We should put a premium on producing fewer, but higher-quality pieces of work and embrace the notion that it will be better for science to produce one solid brick than 10 loose ones. Likewise, the adage that academic hiring committees "can count but cannot read" needs to be banished and replaced with a system the favours a different type of science, such as the one adopted by Ghent University, which has embedded open science best practice as a central organising principle across the institution (https://tinyurl.com/yj78deh7).

On a superficial level, the best possible outcome of taking such an approach is considerably less impressive than we recently thought was possible. An entire career will cover less ground and claim to understand less than we thought was reasonable only a few years ago (Vazire, 2018). And that is precisely how it should be because we have underestimated the challenge that psychological and brain science presents and overestimated our abilities. All things being equal, the result will be slower but firmer science, one that leads more directly to a steady accumulation of knowledge and cumulative science (Frith, 2020; Krakauer, 2019).

3.5. Recognise the value of nonconfirmatory research and calibrate conclusions accordingly

As I outlined earlier in the paper, productive and resource-efficient science involves more than pre-registered, high-powered, confirmatory research. As long as conclusions are appropriately labelled as suggestive, exploratory research is an essential and valuable step in building towards confirmatory research (Tong, 2019). There are also a host of other valuable non-confirmatory steps that should receive more attention before hypotheses are tested, such as forming concepts, developing valid measures, establishing causal relationships between concepts, setting boundary conditions and specifying auxiliary assumptions (Scheel

et al., 2020). Moreover, the division between exploratory and confirmatory research is rarely clear-cut or all-or-none. Most studies typically fall on a continuum between the two poles and often involve both components to some degree. To muddy the water further, what constitutes exploratory and confirmatory research may not be universally agreed upon, and there are also those who propose abolishing the distinction in favour of alternative ways to build theory (Szollosi & Donkin, 2021). Nevertheless, based on existing incentives, it is likely that much of our research practice is already less confirmatory than we acknowledge explicitly (Tong, 2019). Therefore, one way to embrace modesty would be to engage in more nonconfirmatory research because of its inherent value, explicitly label it as such and downregulate conclusions accordingly.

I would also add that much more psychological and brain science research could be purely theoretical rather than experimental, just like occurs in physics where entire subfields are dedicated to theoretical physics alone. Therefore, we should expect, and even demand, that a diverse and interacting set of research approaches will be required to make firmer progress understanding cognition and brain function, rather than focus the majority of our efforts squeezing the life out of the conventional empirical paper. If we fail to diversify programmes of research, we may continue to expect too much from a single piece of research, no matter how well it is designed. As such, we may need to spend more time rebalancing expectations for what even the best empirical work can deliver on its own.

4. Common misconceptions levelled at calls for greater modesty

4.1. Misconception 1: Modesty involves replacing innovation with tedium.

The first misconception equates modesty with the position that innovative work is no longer possible and only "boring" work can be completed. It is true that the last 5-10 years has raised awareness that comprehensive and innovative work is tremendously difficult to do. We cannot continue to make far-reaching claims with such conviction based on the same type of evidence. We need to make more modest conclusions across the board, irrespective of the approach. And stronger conclusions require considerably more resources and effort. Acknowledging this situation does not necessitate that research has to only be "boring" and incremental; it simply means that conclusions should be appropriately re-calibrated to reflect a much tighter link between evidence and inference.

4.2. Misconception 2: Modesty represents a lack of ambition.

The second misconception considers that modesty necessarily reflects a lack of ambition. In fact, one could argue the exact opposite. A modest outlook could be a more ambitious position to take because it will ultimately lead to a greater accumulation of knowledge and minimise the misdirection of scientific funds that overstated claims frequently attract (Frith, 2020; Krakauer, 2019; Vazire, 2018). In short, whether modest or not, the work that one produces can be more or less ambitious depending on one's combined assessment of data quality and value.

4.3. Misconception 3: Modesty provides an excuse for lower standards of research.

The third misconception is that modesty condones poorer or flawed research practices. As I outlined earlier, this is categorically and resolutely not the case. Constructive, technical and detailed criticism remains as fair game as ever. Adopting a more modest approach is not a "prizes for all" proposal. It is just that the inferences that researchers want to make should be placed in the broader context outlined above whereby the difference between data quality and value judgments are made clear. Such justifications should then be interrogated and reviewed like any other.

4.4. Misconception 4: Modesty prevents passionate competition between rival theories.

A call for more modest and proportional inferences, does not prevent research groups from articulating combative counter positions to each other's claims. For example, one may set up purposely adversarial collaborations (Ellemers et al., 2020; Kahneman, 2003), which pre-register relevant terms of engagement, such as what would constitute a "win" for one side over the other. Such an approach may allow more freedom to adopt more extreme positions that usual, under the clear remit that the exercise required an adversarial approach. With this said, the overall outcome of the collaboration would hopefully still be a judgment that required a proportional link between data and inference, which makes it consistent with my overall appeal for modesty. In other words, unless the adversarial collaboration involves an unusually large increase in scale that spans many different methods, approaches, samples etc., inferences should remain fairly modest.

5. Genuine dangers with embracing modesty

There are at least two genuine concerns that come with the proposed shift towards greater intellectual humility. First, modest claims that are presented in old currency terms and embedded within a system that demands exaggeration, may not be well received by all. In fact, the cacophony of "this does not represent a big enough advance" would be deafening from reviewers, editors, grant panels and hiring committees alike. Much like efforts to embrace methodological reform, this is not a trivial issue, which comes with clear costs as well as benefits (Allen & Mehler, 2019; Bielczyk et al., 2020; McKiernan et al., 2016; Poldrack, 2019). As such, it should be a concern for anyone striving to build a career in science, and especially early career researchers who are taking their first steps on the academic ladder.

One way to tackle this challenge, in my view, is to explicitly justify the approach taken. Avoid presenting modest claims in old terms and instead present them within a different, clearly articulated and newer context. Emphasise the value of limiting claims to a, b and c, rather than making the grander claims of d, e and f. By directly comparing a deflationary approach with the available alternatives, it under-

scores the value of the chosen approach. It also makes clear that to evaluate research output, the quality of data needs considering within the context of the aims and values that have been placed on running the project in the first place. This does not guarantee that anyone else has to think that your work is interesting, valuable or well-executed; it just means that a reader understands why you consider grander claims, at this point in time, an unjust overreach that is detrimental to scientific progress.

Moreover, just like any new approach, such as a new approach to statistics, clear justification is essential and has to come front and centre (Cumming, 2012, 2014). For example, one cannot stop using a near-universal approach to inferential statistics in behavioural science, such as the reporting of p-values, without clear justification of the benefits of doing so (Cumming, 2012, 2014). Likewise, one cannot start making narrower claims that purport to make more incremental steps forward without justifying why it is important to do so: because it actually gets the research community closer to our aim, which is to build a cumulative science of mind and brain.

The second concern relates to incentive structures in modern science. Given that incentive structures for publishing in top journals, as well as across hiring committees and grant review panels are dominated by a demand for novel, large-sounding and ultimately exaggerated claims, one might be "pissing in the wind" by trying to take a considerably more deflationary position. In other words, it may be futile to even try a more modest approach, given the powerful and stubborn nature of system-wide incentive structures that are inconsistent with modesty. Incentives are, of course, very important and nearly always misaligned with the production of a cumulative science (Ritchie, 2020). However, as I argued earlier, my view would be that the scale of reform needed requires reform at every conceivable level. And individuals are certainly part of it and should therefore shoulder some of the responsibility, as has been argued previously (Ritchie, 2020; Yarkoni, 2018). This could take the form of more modest initial claims, as well as more routine engagement with public self-correction of previously published work (Rohrer et al., 2021). A hopeful note, which may provide some motivation for a modest approach, would be that real change can only occur when many small steps are taken at many different levels. I am aware, however, that this may simply be a hopelessly naïve perspective.

6. Summary

The replication crisis has taught us that we need to become more modest in our assertions and to steer clear of confident proclamations based on isolated positive results. (Lilienfeld, 2017, p. 663)

To be clear, I do not call for greater modesty from a position of particular strength and certainly do not suggest that my prior research is flawless in this, or any other, regard. I am the product of the same system that I am criticising. I write this from a lived experience of academically growing up in an environment where the overwhelming norm is that big claims can be licensed on small, narrow and inadequate datasets/approaches. Moreover, such claims are heavily rewarded by the scientific system at large, where novelty is favoured at almost all costs. In terms of the reform agenda, I am not favouring one approach over another as the relevant or most important type of reform; instead, without clear evidence otherwise, pluralism in science should dominate. Indeed, one major benefit of science is its broad and diverse base.

I am, however, stating that *all* approaches are limited, and the field would make firmer progress if we did a better job of acknowledging the narrow nature of the evidence that claims are often based upon (Yarkoni, 2020). Indeed, we should avoid the Cult of the Isolated Study (Nelder, 1986; Tong, 2019), and instead stress that an initial or solitary piece of evidence is suggestive and what more comprehensive evidence *would* look like. Of course, one detailed piece of work on any one level can be extremely valuable and worthwhile. It is not a requirement that every piece of work is multi-levelled and multi-method nor is it practically feasible. Rather, I am arguing that it is more important to clearly and explicitly label limitations of each and every approach, whilst also developing means for synthesis across diverse levels of analysis. Such a focus would minimise the perverse incentive to inflate claims and diminish prior/alternative work in order to artificially boost the richness of one's own work. The result would be a firmer footing to build cumulative knowledge.

Finally, it is worth remembering that it is not an idle threat that we may be wasting our time (Yarkoni, 2020): at the present moment, it remains fundamentally difficult for humans to understand the mind and brain, at the level at which we wish to understand it. Nonetheless, I suggest that we do not give up trying quite yet. Instead, we should pivot away from making exaggerated claims and move towards a focus on the limitations of our individual efforts, as well as how we can work together to build ways of thinking that enable integration and synthesis across multiple modalities and levels of description. Will a dose of modesty solve the inherent difficulty of building a cumulative science of psychological and brain function? No. Will it help a little bit? I think so, but I am far from certain.

Acknowledgements

REFERENCES

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *Plos Biology*, *17*(5). https://doi.org/10.1371/journal.pbio.3000246

Apšvalka, D., Cross, E. S., & Ramsey, R. (2018). Observing action sequences elicits sequence-specific neural representations in frontoparietal brain regions. *Journal of Neuroscience*, *38*(47), 10114–10128.

Babbage, C. (1830). *Reflections on the decline of science in England, and on some of its causes*. B. Fellowes.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in Cognitive Sciences*, *15*(5), 200–209. https://doi.org/10.1016/j.tics.2011.03.006

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 48. https://doi.org/10.18637/jss.v067.i01

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., … Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z

Berkman, E. T., & Wilson, S. M. (2021). So Useful as a Good Theory? The Practicality Crisis in (Social) Psychological Theory. *Perspectives on Psychological Science*, 1745691620969650. https://doi.org/10.1177/1745691620969650

Bielczyk, N. Z., Ando, A., Badhwar, A., Caldinelli, C., Gao, M., Haugg, A., Hernandez, L. M., Ito, K. L., Kessler, D., Lurie, D., Makary, M. M., Nikolaidis, A., Veldsman, M., Allen, C., Bankston, A., Bottenhorn, K. L., Braukmann, R., Calhoun, V., Cheplygina, V., … Zhou, X. (2020). Effective Self-Management for Early Career Researchers in the Natural and Life Sciences. *Neuron*, *106*(2), 212–217. https://doi.org/10.1016/j.neuron.2020.03.015

Bommarito, N. (2018). Modesty and Humility. In E. N. Z (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition).

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science*, 1745691620969647. https://doi.org/10.1177/1745691620969647

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Burton, N. (2018). What's the difference between modesty and humility? *Psychology Today*.

Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex*, *49*(3), 609–610.

Chambers, C. D. (2017). *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*. Princeton University Press.

Cobb, M. (2020). *The idea of the brain: a history*. Profile Books.

Cross, E. S., Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage*, *31*(3), 1257–1267. https://doi.org/10.1016/j.neuroimage.2006.01.033

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* Routledge.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Davis, E. (2017). *Post-Truth: Why We Have Reached Peak Bullshit and What We Can Do About It*. Little, Brown Book Group.

Debrouwere, S. (2020). *The conceptual, cunning and conclusive experiment in psychology*.

Ellemers, N., Fiske, S. T., Abele, A. E., Koch, A., & Yzerbyt, V. (2020). Adversarial alignment enables competing models to engage in cooperative theory building toward cumulative science. *Proceedings of the National Academy of Sciences*, *117*(14), 7561–7567. https://doi.org/10.1073/pnas.1906720117

Field, S. A., Tyre, A. J., Jonzén, N., Rhodes, J. R., & Possingham, H. P. (2004). Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters*, *7*(8), 669–675.

Firestein, S. (2012). *Ignorance: How it drives science*. Oxford University Press.

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288.

Fried, E. I., & Robinaugh, D. J. (2020). *Systems all the way down: embracing complexity in mental health research*. Springer.

Frith, U. (2003). *Autism: Explaining the Enigma*. Blackwell Publishing.

Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, *24*(1), 1–2.

Gazzaniga, M. S. (2010). Neuroscience and the correct level of explanation for understanding mind: An extraterrestrial roams through some neuroscience laboratories and concludes earthlings are not grasping how best to understand the mind-brain interface. *Trends in Cognitive Sciences*, *14*(7), 291–292. http://www.sciencedirect.com/science/article/B6VH 9-504K881-1/2/5a8fe340650d1f7dcf519103183a0dcc

Gervais, W. M. (2021). Practical Methodological Reform Needs Good Theory. *Perspectives on Psychological Science*, 1745691620977471. https://do i.org/10.1177/1745691620977471

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218.

Gleick, J. (2004). *Isaac Newton*. Vintage.

Gleick, J. (2011). *Chaos: Making a new science*. Open Road Media.

Gray, K. (2017). How to Map Theory: Reliable Methods Are Fruitless Without Rigorous Theory. *Perspect Psychol Sci*, *12*(5), 731–741. https://doi.org/1 0.1177/1745691617691949

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, 1745691620970585. https://doi.org/10.1177/17456916 20970585

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. MIT Press.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nat Neurosci*, *9*(10), 1218–1220. https://doi.org/10.1038/n n1770

Herrero, L. (2014). *'Simplify, then exaggerate': a saying from journalism, with different versions in our organizations*. https://leandroherrero.com/first-simpli fy-then-exaggerate-a-journalism-saying-with-differe nt-versions-in-our-organizations/

Hintzman, D. L. (1991). Why are formal models useful in psychology? In *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock.* (pp. 39–56). Lawrence Erlbaum Associates, Inc.

Hoekstra, R., & Vazire, S. (2020). *Intellectual humility is central to science*. https://psyarxiv.com/edh2s/

Huber, D. E., Potter, K. W., & Huszar, L. D. (2019). Less "story" and more "reliability" in cognitive neuroscience. *Cortex*, *113*, 347–349. https://doi.org/1 0.1016/j.cortex.2018.10.030

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, *58*(9), 723.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Galliani, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152. http s://doi.org/10.1027/1864-9335/a000178

Krakauer, J. (2019). *John Krakauer/Interviewer: Neuron*. Neuron, CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA.

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial Introduction with R*. Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev*, *25*(1), 178–206. http s://doi.org/10.3758/s13423-016-1221-4

Lakens, D. (2020). Review of "The Generalizability Crisis" by Tal Yarkoni. *The 20% Statistician*.

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S.-C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., … Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, *2*(3), 168–171. https://doi.org/10.1038/s41562-018-0311-x

Lean, M. E. J., Astrup, A., & Roberts, S. B. (2018). Making progress on the global crisis of obesity and weight management. *BMJ*, *361*, k2538.

Lilienfeld, S. O. (2017). Psychology's Replication Crisis and the Grant Culture: Righting the Ship. *Perspectives on Psychological Science*, *12*(4), 660–664. https://doi.org/10.1177/1745691616687745

Lilienfeld, S. O., Lynn, S. J., O'Donohue, W. T., & Latzman, R. D. (2017). Epistemic humility: An overarching educational philosophy for clinical psychology programs. *Clinical Psychologist*, *70*, 6–14.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

McGinn, C. (1989). Can We Solve the Mind--Body Problem? *Mind*, *98*(391), 349–366.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

Meichenbaum, D., & Lilienfeld, S. O. (2018). How to spot hype in the field of psychotherapy: A 19-item checklist. *Professional Psychology: Research and Practice*, *49*(1), 22–30. https://doi.org/10.1037/pro0000172

Morton, J. M. (2004). *Understanding developmental disorders: a causal modelling approach*. Blackwell.

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., … Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. https://doi.org/10.1038/s41562-016-0021

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Navarro, D. J. (2019). Between the Devil and the Deep Blue Sea: Tensions Between Scientific Judgement and Statistical Model Selection. *Computational Brain & Behavior*, *2*(1), 28–34. https://doi.org/10.1007/s42113-018-0019-z

Navarro, D. J. (2021). If Mathematical Psychology Did Not Exist We Might Need to Invent It: A Comment on Theory Building in Psychology. *Perspectives on Psychological Science*, 1745691620974769. https://doi.org/10.1177/1745691620974769

Nelder, J. A. (1986). Statistics, science and technology. *Journal of the Royal Statistical Society: Series A (General)*, *149*(2), 109–121.

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annu Rev Psychol*, *69*, 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Neyman, J., & Pearson, E. S. (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*(694–706), 289–337.

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van't Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*, *45*, 137–141.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Oster, E. (2013). *Expecting Better: Why the Conventional Pregnancy Wisdom is Wrong--and what You Really Need to Know*. Penguin.

Plomin, R. (2018). *Blueprint: How DNA Makes Us Who We Are*. Allen Lane.

Poldrack, R. A. (2019). The costs of reproducibility. *Neuron*, *101*(1), 11–14.

Ramsey, R. (2020). Advocating for the credibility revolution. *Cognitive Psychology Bulletin*, *5*. https://psyarxiv.com/3kwnu

Rawlins, M. D., & Culyer, A. J. (2004). National Institute for Clinical Excellence and its value judgments. *BMJ*, *329*(7459), 224–227.

Raynor, H. A., & Champagne, C. M. (2016). Position of the Academy of Nutrition and Dietetics: interventions for the treatment of overweight and obesity in adults. *Journal of the Academy of Nutrition and Dietetics*, *116*(1), 129–147.

Reichle, E. D. (2020). *Computational models of reading: A handbook*. Oxford University Press.

Ritchie, S. (2020). *Science fictions: Exposing fraud, bias, negligence and hype in science*. Random House.

Roberts, R., & Wood, J. (2003). Humility and Epistemic Goods. In M. DePaul & L. Zagzebski (Eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (pp. 257–279). Clarendon Press.

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction. *Perspectives on Psychological Science*, 1745691620974697. https://doi.org/10.1177/1745691620974697

Rohrer, J. M., Tierney, W., Uhlmann, E. L., DeBruine, L. M., Heyman, T., Jones, B., Schmukle, S. C., Silberzahn, R., Willén, R. M., Carlsson, R., Lucas, R. E., Strand, J., Vazire, S., Witt, J. K., Zentall, T. R., Chabris, C. F., & Yarkoni, T. (2021). Putting the Self in Self-Correction: Findings From the Loss-of-Confidence Project. *Perspectives on Psychological Science*, 1745691620964106. https://doi.org/10.1177/1745691620964106

Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology*, *21*(4), 308–320.

Rubin, M. (2019). The costs of HARKing. *The British Journal for the Philosophy of Science*.

Sanbonmatsu, D. M., & Johnston, W. A. (2019). Redefining Science: The Impact of Complexity on Theory Development in Social and Behavioral Research. *Perspectives on Psychological Science*, *14*(4), 672–690. https://doi.org/10.1177/1745691619848688

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2020). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, 1745691620966795. https://doi.org/10.1177/1745691620966795

Shiffrin, R. M., Börner, K., & Stigler, S. M. (2018). Scientific progress despite irreproducibility: A seeming paradox. *Proceedings of the National Academy of Sciences*, *115*(11), 2632–2639. https://doi.org/10.1073/pnas.1711786114

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational Social Psychology* (pp. 311–331). Routledge.

Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*(4), 207–218. https://doi.org/10.1027/1864-9335/a000425

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384. https://doi.org/10.1098/rsos.160384

Spiegelhalter, D. (2019a). *The Art of Statistics: Learning from Data*. Penguin.

Spiegelhalter, D. (2019b). *What's behind the sausage wars? Three questions to ask of any contested claim*. https://medium.com/wintoncentre/are-we-individuals-or-members-of-populations-the-deeper-issues-behind-the-sausage-wars-a067aebf2063

Szollosi, A., & Donkin, C. (2021). Arrested Theory Development: The Misguided Distinction Between Exploratory and Confirmatory Research. *Perspectives on Psychological Science*, 1745691620966796. https://doi.org/10.1177/1745691620966796

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2019). Is preregistration worthwhile. *Trends in Cognitive Sciences*, *24*(2), 94–95.

Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science. *The American Statistician*, *73*(sup1), 246–261. https://doi.org/10.1080/00031305.2018.1518264

van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, 1745691620970604. https://doi.org/10.1177/1745691620970604

Vazire, S. (2018). Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2015). Intellectual Humility: Owning Our Limitations. *Philosophy and Phenomenological Research*, *91*(1), 1–31.

Yarkoni, T. (2018). *No, It's Not The Incentives - It's You*. https://www.talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/

Yarkoni, T. (2020). The generalizability crisis. *Behav Brain Sci*, 1–37. https://doi.org/10.1017/s0140525x2000 01685

SUPPLEMENTARY MATERIALS

Peer review history

Download: https://collabra.scholasticahq.com/article/24091-a-call-for-greater-modesty-in-psychology-and-cognitive-neuroscience/attachment/61377.docx?auth_token=LH-pAXs0mA4WYy0j1ip1