Computational Brain & Behavior
https://doi.org/10.1007/s42113-019-00036-z

ORIGINAL PAPER

Cultural Problems Cannot Be Solved with Technical Solutions Alone

Simon D. Lilburn[1] · Daniel R. Little[1] · Adam F. Osth[1] · Philip L. Smith[1]

Abstract

A crisis in psychology has provoked researchers to seek remedies for bad practices that might damage the integrity of the discipline as a whole. The ardor for wholesale reform has led to a suite of proposed technical solutions, some of which are considered in the context of computational modeling by the target article. Any technical solution, however, must be placed within a larger cultural and scientific context to be effective (or, indeed, meaningful at all). Many of the suggestions presented in the target article represent good practice in computational cognitive modeling but, even then, still require some amount of nuance in the consideration of the relationship between practice and theory. We consider two examples—model preregistration and bookending—as a means of examining the limits of any proposed technical solution.

Q2 **Keywords**

> In moments of crisis, people are willing to hand over a great deal of power to anyone who claims to have a magic cure… — Naomi Klein, *The Shock Doctrine* (2007)

In 2010, the year before the allegations against Diederik Stapel surfaced (see the report by the Levelt et al. 2012) and the publication of Bem's *JPSP* article on the putative existence of ESP (Bem 2011), popular science writer Jonah Lehrer (2010) noted a worrying trend for apparently real effects in psychology and elsewhere to evaporate with time. This observation was followed by specific failures to replicate effects within social priming (Harris et al. 2013; Shanks et al. 2013) and, then, much more general failures to replicate effects across psychology as a whole. The Open Science Collaboration (2015) found, in line with Lehrer's observation, that the number of statistically significant results fell from 97 to 36% upon replication.[1]

That cluster of apparently negative outcomes—and similar incidents in the intervening period—has led to a flurry of commentary and introspection attempting to anatomize and remedy the set of problems. Many of the potential solutions focus on introducing mechanisms of additional conformity at a technical level: among them, a revolution of widespread preregistration (Nosek et al. 2018) or merit badges (Kidwell et al. 2016) or making null hypothesis significance testing uniformly more conservative (Benjamin et al. 2018).

In the current issue, the target article by Lee and colleagues (Lee et al. in press) provides a number of technical solutions as they may pertain specifically to computational and mathematical modeling within psychology. We agree Q3 that each of the proposed measures has clear utility, when applied appropriately and thoughtfully, but also wish to express reservations about the general purpose nature of such solutions, particularly any solution pitched at a purely technical level. The solutions proposed in the target article include the use of a combination of qualitative and quantitative benchmarks against which to test models, the use of multiple data sets to test a model's generalizability, the use of multiple fit indices to ensure the robustness of conclusions, and the use of cross-validation and out-of-sample prediction to guard against overfitting. Different subsets of these techniques would be understood by most researchers as hallmarks of good modeling in psychology as it is currently practiced (see, for instance, Heathcote et al. 2015) but each represents an intervention in the technical side of scientific practice that may collectively lead to improvements

[1] Ironically, Lehrer himself later resigned from the *New Yorker* after fabricating Bob Dylan quotes to support the argument of his book *Imagine*, lending some support towards the idea that, whatever cultural issue that causes novelty to be preferenced over evidence is at play, it is not solely located within science, let alone psychology.

✉ Simon D. Lilburn
lilburns@unimelb.edu.au

Q1 [1] Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria, Australia

in the way inferences are drawn from data, but leave the underlying cultural issues unaddressed.

Advocacy, by its general nature, can be one-sided in its presentation: it operates from putting forth a strong position in response to a perceived problem, but often must leave any considerations which might moderate that position and might establish parameters on the candidate solution to others. This becomes a problem when these solutions are actually applied in practice, and becomes a particular problem when that advocacy is invoked in tandem with a singular conception of what constitutes "good science" or "ideal practice." To offer an example: the broad catch-cry of "increase your sample size," reasonable when "sample size" is interpreted as "the number of observations" available, was interpreted by some—largely in the review of psychophysical papers—as being confined to increasing the total number of individual participants, which led to the article by Smith and Little (2018). The transformation of a very minimal good faith prescription (collect more data!) into an injunction (small *N* research is not worth publishing) reflects a cultural problem that has remained intact despite efforts to frame the problem in technical terms, namely in terms of reducing the number of underpowered studies.

What is true for sample size is likely also true for other kinds of methodological advocacy, such as the target article's advocacy of a sharper distinction between exploratory and confirmatory modeling studies. Like prescriptions about sample size, proposals of this nature tend to become divorced from any qualifications and caveats that accompanied them when they were first made, through no fault of the advocates themselves. This is especially so when attempts are made to codify a given set of ideas about good practice within the editorial and reviewing standards of journals. While well-intended, these attempts can result in a set of restrictive prescriptions that frustrate the very ends they were seeking to promote.

In this response, we wish to discuss two proposals of the target article, model preregistration and bookending, as solutions that can be powerful but do not sidestep the consideration and judgement of scientists operating in good faith and must always be subordinated to the question of whether the procedure naturally aligns with the question being asked.

Preregistration, or, the Procrustean Bargain

The interplay between model construction and testing against data is explored in the target article through the lens of preregistration, registered model reports, and "blind modeling."[2] These concepts, and particularly modeling preregistration, import the larger conversation within psychology about the potential benefits of widespread preregistration (Nosek et al. 2018) into a computational modeling framework. Without doubt, preregistration is a useful methodological tool. Its efficacy in some settings is unquestionable: Kaplan and Irvin (2015) noted a large change in the number of positive results reported in randomized clinical trials from before and after the year 2000 with the introduction of preregistration at ClinicalTrials.gov, going from 57% of studies with non-null effects prior to 2000 to only 8% afterwards.

This example, however, also highlights some of the limits of preregistration when taken as a general normative model for science. A clinical trial is undertaken when the substantive science is complete: it would be worrying to learn something dramatically new about an intervention being administered to human subjects at the point of a clinical trial. That regime also tends to be very well suited to null hypothesis significance testing, in the sense that intervention efficacy, rather than process, is all that matters. The actual execution of any preregistered plan should be, in the best case, "dumb," in that running a preregistered study should resolve to the rote implementation of a fully specified procedure in a way that does not require the intervention of a scientist, but can be handled by any technician (or, better, an automated script on a computer).

Most science, and most modeling in particular, does not readily fit this frame. Aside from proven cases, such as clinical trials, or areas with a known propensity of questionable practices—where "registered reports" provide an appropriate way to enforce a methodological "firewall" between initial assumptions and final results—most science, and most modeling in particular, requires scientific judgement. Even rigorous, well-specified models are confronted with ancillary assumptions or unanticipated design questions. Some of these assumptions are interesting, in that they may lead to a theoretically useful insight, and some are

[2] We think postregistration is a useful idea. Many modelers will have been frustrated by editors asking them to remove material detailing the full range of model variants they considered because it is seen as dry and indigestible. But we worry that the effort required to do justice to postregistration as envisaged in the target article means it is an idea that will be honored mainly in the breach. Postregistration is tantamount to the requirement that modeling studies be accompanied by substantial supplementary materials sections. Laboratory notebooks are usually aides memoire for researchers rather than public records intended to communicate to others. The effort required by authors, reviewers, and editors to turn them into truly useful adjuncts to scientific practice should not be underestimated.

just lapses in foresight. Among the discourse of those advocating for preregistration as a general movement, however, the difference between "predictions" in the mode of, say, cross-validation (i.e., a single data format is obtained and a single set of algorithms are used for producing predictions) and "prediction" in a broader sense of substantive theoretical consequences (e.g., relativity predicting gravitational waves even though the design and operation of LIGO would have been entirely unknown to Einstein) is somewhat elided. That distinction is, however, critical: one type of prediction is rote to implement, obviates any need for the scientist to intervene, and fits neatly within a totally prespecified framework, and the other requires substantial theoretical work that would be difficult—if not impossible—to fully anticipate.

Indeed, even in cases where an apparently "confirmatory" frame is used, preregistration may not add much more over sufficiently well-developed model comparisons. Consider two recent and (to our minds) theoretically interesting cases: first, the possible existence of "collapsing boundaries" in sequential-sampling models, where the response thresholds change as a function of time, and, second, the underlying structure of representations within visual working memory. In both cases, the literature as a whole converged upon a particular set of candidate models where distinct predictions for certain tasks could be made.

On the question of collapsing boundaries, Hawkins et al. (2015) used nine different datasets, from humans and nonhuman primates, to indicate that the participant populations and the way in which the studies were conducted had a strong effect on whether boundaries appeared to collapse or not. They found that the majority of human participants exhibited strong support for the standard diffusion model with an unchanging response boundaries, judged using an approximation to posterior model probabilities based upon the BIC, but it may be affected by practice or the way that rewards are administered. On the question of visual working memory structure, van den Berg et al. (2014) formulated a variety of different models that combined alternative assumptions about the structure and capacity of memory and examined data from ten published continuous report experiments from different laboratories. They found that the best model—judged in terms of AIC—was one in which the quality and number of items within memory varied across trials, but that the memory system did not itself have an item-capacity limit.

Because the class of scientifically interesting models is not closed, neither of these studies represents the last word on the questions they sought to address, but both made important contributions to their respective literatures. Neither study was preregistered and, to our way of thinking, neither suffered as a result. In either instance, the scientific credibility of the study is inherent in its own internal logic, with the close correspondence between theory and analysis, rather than whether the procedures that were used and the alternatives that were considered were graven in stone prior to seeing the data. In both studies, the researchers laid out a set of alternative models that made differing predictions in relation to the overarching research question and then established the generality of the finding by applying the models to a number of different data sets. Importantly, both studies addressed the issues of replicability and generality as a natural part of ordinary scientific practice. Contrary to the views of the preregistration movement, it makes no sense to dismiss the studies prior to these as "merely exploratory," and to accord these studies special status as "confirmatory" (a feeling conveyed by, e.g., Wagenmakers et al. 2012). Rather the confirmation focus grows organically out of the exploration focus at a time when competing claims emerge that need to be adjudicated. No reform is needed because the ordinary practice of scientific inquiry and debate naturally concentrates the research effort in the places where it is required. In both these cases, the literature as a whole had converged upon a set of competing models that had been supported by previous research. The studies sought to adjudicate between the competing models in a systematic way by applying them to multiple data sets collected under a variety of different conditions by different investigators in different laboratories.

Other suggestions in the target article, like blinding, work to emphasize this dichotomy between exploration and confirmation, and imply that confirmation is somehow preferable. But good science, and good current modeling practice, involves an organic mix of exploration and confirmation—a mix that resists categorization as wholly one or the other. When these practices form some sort of normative framework of scientific inquiry—however minimal that framework might appear—questions about how to thoughtfully diverge from that dichotomy or how to pursue other means of scientific progress become much more difficult to prosecute.

Bookending, or, Judging Model Fits by Their Cover

In providing a definition of "bookending," the authors of the target article preface their discussion with a comment that model comparison is inherently relative. This is true, but in a trivial way: if model comparison is the act of comparing two or more models, then it must be, by definition, relative. However, construed slightly more broadly, one could say that model comparison also encompasses the act of comparing the model to the data in some sort of absolute sense: Does the model provide an adequate sense of the data obtained?

Typically, the difficulties involved in assessing absolute fit arise because of the presence in data of theoretically unimportant, unmodeled sources of variability, such as those that violate the independent, identically distributed random variables assumption on which classical likelihood-based statistics depend. However, the solution to these difficulties is not to recast the problem as one of assessing relative rather than absolute fit. When using routine procedures to assess the relative performance of members of a set of candidate models, it is very easy to lose sight of the more pertinent question of absolute model performance.[3]

Both relative and absolute model performance require a judgement call about which variance in the experimental data is important. An experiment in which the precession of a top on desk is measured will not fully conform to the calculations of force computed from a Newtonian analysis of gravity and the angular momentum directly applied to the top; an experiment in which the precession of the perihelion of mercury is measured will not fully conform to calculations of force computed from a Newtonian analysis of gravity and the angular momentum applied directly to the planet. Unfortunately, the apparently parallel failure of theory in those cases reflects wholly different outcomes for science. Adequately explaining the latter case involves a consideration of relativity, and this consideration reformed our understanding of the universe; adequately explaining the former case might involve a consideration of air drag or thermal convection, and this consideration would reform our understanding of the air conditioner in the room. When none of the models being compared fit the data in an absolute sense (as is the case in the authors' Figure 2), then the "meaning" of the errors remains unknown: i.e., is the failure of the model due to the universal gravitational constant or the air conditioning?

That parts of observed variance are not equal in terms of scientific interest is no revelation to people who deal closely with specifying and testing models, but it does hamper any stronger normative expectations about how science ought to conduct itself. Even the formulation of a null model beyond simple standard sampling distributions—which are often maligned as being radically implausible in psychology (as argued by, e.g., Meehl 1990, 1997; Lykken 1968)—requires consideration about the interesting and uninteresting components of variation. The solution to this problem is exactly the kind of informal and distributed negotiation that occurs in science: iterative steps, punctuated by creative leaps, which inform the set of models and variants that scientists consider interesting.

Where some bookends might be readily apparent, others require just as much consideration as what might be considered theoretically substantive models (because, in effect, they are theoretically substantive models) and the process of making them useful often requires the same sort of close consideration of underlying processes and data that makes fully specifying them in advanced difficult in all but a limited number of cases. Stronger versions of Bayesian model averaging and selection which require explicit prior probabilities placed on models themselves raise even more complex issues along these lines: how should one deal not only with the fact that not all models are included within the candidate set (so-called $\mathcal{M}$-open scenarios; see, Bernardo and Smith 1994; Vehtari and Ojanen 2012) but also with the fact that the candidate set might (and probably should) change from the model specification a priori to subsequent observations of data. Even a strong methodological firewall between model specification and data collection is difficult to reconcile with the natural progression of scientifically interesting models.

Truth on the Installment Plan

There is no way, in our mind, to legislate bad science out of existence. Our view is that the issues that plague psychological science at the moment—those outside of simple bad faith fabrication and fraud—are, at a broad level, issues with an incentive structure that pushes individuals to produce substantive conclusions where none are licensed.[4] The problem with this scenario is not, generally speaking, the statistics themselves, but the gap between the statistics, the properties of measurement, and the mental model within the researcher about how these two things relate (Smith and Little 2018).

As such, the success or failure of any intervention designed to make science "good" or "robust" is not necessarily located within the technical or methodological rigor of that intervention, but rather the relationship between the types of questions researchers wish to ask and answer, and the tools they have at their disposal. Our sense is that technical interventions are not "good" or "robust" in isolation, but

[3]This is a point made beautifully by Navarro (2019) within this journal.

[4]In making our counterargument, we do not seek to provide cover for bad actors, but rather, think that it is just as likely that those acting in bad faith will find a way to game the preregistration system, potentially by making incomplete or late preregistrations, just as they have the null hypothesis significance testing one. The solution is as it has always been: skepticism, peer review, and due diligence in examining published claims. Modeling helps in this endeavor because, in most cases, the outcome of a model is easily reproduced by pushing a button on a computer. This is naturally why we are in full agreement with authors' desire to promote openness and sharing of materials and code.

are useful only in the sense that they align well, that is to say naturally, with the questions that researchers wish to answer.[5] Modeling often has the sense of jointly discovering the question when trying to formulate the answer: it is a creative act and one that does not often neatly fall into the category of "exploration" or "confirmation." Consider, for instance, the work of Marr (1982), who detailed a computational theory of how three-dimensional representations might be derived from a retinal image, or Anderson (1990), who examined cognitive phenomena such as categorization and memory by proposing behavioral functions that would provide the optimal solution in a given environment to some larger adaptive goal. These represent some of the key contributions made within psychology as works of theory-building, and works in which the labels "exploration" or "confirmation" do not apply in any meaningful way. The need to force research into categories in which the intentions of the researcher do not neatly align with the method provides the outline for the next crisis.

The creative aspect of science *is* science. When Rutherford said, (in)famously, that all science outside of physics was "stamp-collecting," he painted a world in which all the useful, creative, vital enterprise of science was the privilege of physicists, and derogated everything else as just confirmatory box ticking. When students learn, in their advanced undergraduate years, probability modeling outside of the null hypothesis significance testing framework, we often witness their relief in realizing that there is an escape from a type of conformity which aims to put the richness of their experimental and theoretical ideas into molds well suited to split-plot agronomy or clinical trials. That introduction to modeling provides the first look at a larger language with which they can more clearly, more neatly express their theoretical ideas. That language of expression and the ways it can be applied has limitations, to be sure, but the closer alignment between theoretical ideas and their formal expression leads to a stronger coupling between intention and practice, and a sharper distinction where the two diverge.

[5]This is also one reason why we cannot, despite all of the limitations of null hypothesis significance testing and its applications in the wild, bring ourselves to endorse abandoning statistical significance entirely. Although it was never intended as such, we could see the wholesale abrogation of a type of statistical inference, either as a piece of advice for the field or as a directive at the journal-level, as tending in the direction of the same all-or-none thinking as individual scientists misusing statistical significance testing in the first place: the kind of cargo cult science that inappropriately identifies that a statistical procedure can provide some license to make inferential statements without understanding the mechanism by which it does so. At worst, it leads people away from an opportunity to gain a nuanced understanding of the limitations of any particular approach to statistical inference to once again promote a regime where understanding is unnecessary, only a set of imperatives are worth understanding; simply put, it replaces the cookbook with the rulebook.

The tools offered by the target article, we believe, are useful insofar as they allow researchers more expressive, more precise ways of expressing the ideas they wish to investigate, and more rigorous ways to pursue those ideas. But their utility and applicability is easy to overestimate. No procedure provides privileged access to an otherwise unseen realm of truth. The authors of the target article have been careful in discussing caveats and exceptions in their discussion of suggestions for reform. We hope that authors, reviewers, and journal editors will be open to new tools and techniques, while remaining mindful to the pitfalls of their uncritical application.

Acknowledgments Thank you to David Wakeham, Christina van Heer, David Sewell, Jason Zhou, and Elle Pattenden for their thoughtful comments and questions.

References

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale: L. Erlbaum Associates.

Bem, D.J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. https://doi.org/10.1037/a0021524.

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., Boeck, P.D., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Fehr, E., Fidler, F., Field, A.P., Forster, M., George, E.I., Gonzalez, R., Goodman, S., Green, E., Green, D.P., Greenwald, A.G., Hadfield, J.D., Hedges, L.V., Held, L., Ho, T.H., Hoijtink, H., Hruschka, D.J., Imai, K., Imbens, G., Ioannidis, J.P.A., Jeon, M., Jones, J.H., Kirchler, M., Laibson, D., List, J., Little, R., Lupia, A., Machery, E., Maxwell, S.E., McCarthy, M., Moore, D.A., Morgan, S.L., Munafó, M., Nakagawa, S., Nyhan, B., Parker, T.H., Pericchi, L., Perugini, M., Rouder, J., Rousseau, J., Savalei, V., Schönbrodt, F.D., Sellke, T., Sinclair, B., Tingley, D., Zandt, T.V., Vazire, S., Watts, D.J., Winship, C., Wolpert, R.L., Xie, Y., Young, C., Zinman, J., Johnson, V.E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6. https://doi.org/10.1038/s41562-017-0189-z.

Bernardo, J.M., & Smith, A.F.M. (1994). *Bayesian theory*. New York: Wiley, Chichester, Eng.

Harris, C.R., Coburn, N., Rohrer, D., Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLOS ONE*, *8*(8), e72,467. https://doi.org/10.1371/journal.pone.0072467.

Hawkins, G.E., Forstmann, B.U., Wagenmakers, E.J., Ratcliff, R., Brown, S.D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, *35*(6), 2476–2484. https://doi.org/10.1523/JNEUROSCI.2410-14.2015.

Heathcote, A., Brown, S.D., Wagenmakers, E.J. (2015). An introduction to good practices in cognitive modeling. In *An introduction to model-based cognitive neuroscience* (pp. 25–48): Springer.

Kaplan, R.M., & Irvin, V.L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLOS ONE*, *10*(8), e0132,382. https://doi.org/10.1371/journal.pone.0132382.

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S., Nosek, B.A. (2016). Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), e1002,456. https://doi.org/10.1371/journal.pbio.1002456.

Klein, N. (2007). *The shock doctrine: the rise of disaster capitalism*, 1st. New York: Metropolitan Books/Henry Holt. oCLC: 128236664.

Lee, M.D., Criss, A., Devezer, B., Donkin, C., Etz, A., Leite, F.P., Matzke, D., Rouder, J.N., Trueblood, J.S., Vandekerckhove, J. (in press). Robust modeling in cognitive science. Computational Brain and Behavior.

Q4 Lehrer, J. (2010). The Truth Wears Off. The New Yorker, http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off.

Levelt, W.J.M., Noort, E., Drenth Committees (2012). Flawed science: the fraudulent research practices of social psychologist Diederik Stapel. Tech. rep. Tilburg University, the University of Groningen, and the University of Amsterdam. Q5

Lykken, D.T. (1968). Statistical significance in psychological research. *Psychological bulletin*, *70*(3), 151–159.

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. Cambridge: The MIT Press.

Meehl, P.E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195.

Meehl, P.E. (1997). The problem is epistemology, not statistics: replace significance tests by confidence intervals and quantify accuracy of risky numerical prediction. In L.L. Harlow, S.A. Mulaik, J.H. Steiger (Eds.) *What If There Were No Statistical Tests?, Erlbaum, Mahwah, N.J* (pp. 393–425).

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.

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.

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

Shanks, D.R., Newell, B.R., Lee, E.H., Balakrishnan, D., Ekelund, L., Cenac, Z., Kavvadia, F., Moore, C. (2013). Priming intelligent behavior: an elusive phenomenon. *PLOS ONE*, *8*(4), e56,515. https://doi.org/10.1371/journal.pone.0056515.

Smith, P.L., & Little, D.R. (2018). Small is beautiful: in defense of the small-N design. Psychonomic Bulletin & Review, pp. 1–19, https://doi.org/10.3758/s13423-018-1451-8.

van den Berg, R., Awh, E., Ma, W.J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234.

Vehtari, A., & Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, *6*, 142–228. https://doi.org/10.1214/12-SS102.

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.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES:

Q1. Please check if the affiliation is presented correctly.
Q2. Keywords are desired. Please provide.
Q3. Lee et al. (in press) has been cited in the sentence "In the current issue, the target...." Please check if correct.
Q4. Please check captured year in reference Lee et al. (in press) if correct.
Q5. Please check author names in reference Levelt et al. (2012) if correct.