to imply that practitioners are doing a poor job of communicating I-O psychology science. As Grand et al. suggest, when they do it well, practitioners secure a competitive advantage through their dedication to rigorous science.

Some psychologists have suggested that professional organizations that represent practitioners operate more as guilds than as defenders of robust science. The growth of the Association for Psychological Science (APS) from the American Psychological Association (APA) is evidence of this division. However, I do not believe this has to be such a concern for I-O psychologists. Professional societies such as SIOP could be of service to all its members—scientists and practitioners alike—by more actively recognizing good science, and calling out junk science.

The recommendations Grand et al. (2018) have for improving the robustness of science from an academic point of view are practical and irrefutable. I think a similar level of emphasis can be dedicated to ensuring that the corporate customer is properly educated as well. Professional associations are uniquely positioned to educate consumer audiences with the benefit of unbiased credibility and with members from the scientific community who can responsibly translate science to a practical audience.

Reference

Grand, J. A., Rogelberg, S. G., Allen, T. D., Landis, R. S., Reynolds, D. H., Scott, J. C., . . . Truxillo, D. M. (2018). A systems-based approach to fostering robust science in industrial-organizational psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 11*(1), 4–42.

In Defense of HARKing

Jeffrey B. Vancouver
Ohio University

Science is a complex task. It involves the creation and dissemination of knowledge. The creation of knowledge requires identifying and abstracting patterns (i.e., identifying phenomena and theorizing about the processes that bring it about), as well as systematically observing to better see and quantify the patterns (e.g., effect size estimating) or assess the validity of the abstractions used to explain the patterns (i.e., theory testing). To help (a) hone in on what observations would be useful and (b) communicate what the pat-

Jeffrey B. Vancouver, Department of Psychology, Ohio University.
Correspondence concerning this article should be addressed to Jeffrey B. Vancouver, Ohio University, Department of Psychology, 200 Porter Hall, Athens, OH 45701. E-mail: vancouve@ohio.edu

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

terns mean, we are encouraged to develop and report hypotheses. That is, *strategically*, hypotheses facilitate the planning of data collection by helping the researcher understand what patterns need to be observed to assess the merit of an explanation. Meanwhile, *tactically*, hypotheses help focus the audience on the crucial patterns needed to answer a question or test a theory. When the strategic hypotheses are not supported, it raises a question regarding what to do tactically. Depending on the result (i.e., different direction; null), one might construct a hypothesis to facilitate dissemination without reporting this *post hoc* construction or remove mention of a hypothesis altogether. This practice is called HARKing (i.e., hypothesizing after results are known). HARKing has been so disparaged as to be considered a "detrimental research practice" (Grand et al., 2018, p. 6). As such, the Society for Industrial and Organizational Psychology's (SIOP) Robust and Reliability Science task force appears to be recommending that HARKing not be taught by educators, encouraged by reviewers or editors, or practiced by authors. I do not agree with those recommendations, and I elaborate on my position below.

Background

A review of the current literature exposes mostly articles that condemn HARKing (e.g., Bosco, Aguinis, Field, Pierce, & Dalton, 2016), but it also includes articles that are more nuanced about its merits or demerits (e.g., Hollenbeck & Wright, 2017). This comment is one of the latter type. That is, I do not condone HARKing uniformly. Indeed, I recognize the central argument that HARKing can capitalize on chance. In particular, the concern is that an unexpected significant finding or an unexpected, nonsignificant finding stems from sampling error (i.e., it is local to the dataset). Though true, the probability of capitalizing on chance is moderated by the experiment-wise error rate. That is, the more statistical tests done on a dataset, the more likely a Type-I error will be made (i.e., interpreting an effect as something other than null). Additional concerns such as biases in judgment and concerns about deception in the conduct of science are also associated with HARKing. I will touch on all these issues.

Definitional Issues

HARKing is an acronym for "hypothesizing after results are known," but when Kerr (1998) talked about it, he also advocated the simple solution of noting the order of hypothesizing relative to data collection. This is what Hollenbeck and Wright (2017) call transparent HARKing (i.e., THARKing). Moreover, Kerr did not want to stifle what I call TARKing, which is theorizing after results are known. Thus, the primary issue is lack of transparency, not the hypothesizing after results are known. Thus, HARKing means secretly hypothesizing after results are known, which is what Hollenbeck and

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

Wright (2017) called SHARKing. To be clear, it is stifling this secretive behavior that I am suggesting might be bad for science in some cases.

Before proceeding, we need a bit more on the definition of HARKing. I think the kind most commonly assumed is to add hypotheses not originally proposed. However, removing hypotheses after results are known, secretively, is also HARKing. Personally, I commit the latter more often than the former, though I like to think I do neither all that often. Another potential form of HARKing is modified hypotheses. That is, changing some detail regarding the way the hypothesis is stated. It displeases me how often I do this or encourage this in my students' theses and dissertations given the scrutiny I and the committee presumably give to the proposal. Sometimes these changes need to be transparent and sometimes not, for the same reasons I think adding or subtracting hypotheses sometimes should be transparent or not done. In this way, my point is very similar to Hollenbeck and Wright's (2017), which is that legitimate or even good forms of hypothesizing after results are known are being swept into the HARKing-is-bad net. The demarcation for Hollenbeck and Wright, like Kerr (1998), was whether it was secretive or not. The demarcation for me is, unfortunately, a fuzzier line. It is whether the quality of the logical arguments, evidence, and methods warrant HARKing.

The Roles of Hypothesizing

One of the issues motivating this comment is not so much that I practice HARKing occasionally as an author or encourage it periodically as a reviewer, but that I teach it in my research methods class. That is, I explain that there might be instances where one is faced with the situation where adding, dropping, or revising a hypothesis is advisable after the results are known, and that being transparent about that will likely not serve science. That such behavior might dovetail with one's own interests is irrelevant; I am interested in better science.

In the interest of science as well as the individual researcher, I point out that hypothesizing *a priori* has substantial strategic value because it specifies the tests that will be of diagnostic value to the research question being asked. However, I also point out that most tests are confirmatory. That is, they are of the nature that if theory t is relevant here, prediction p should be observed. This is not very diagnostic if alternative explanations for p are likely. Thus, one should also use the hypotheses to focus attention toward alternative explanations for the predictions so that additional design elements might be included to help assess these alternative explanations. In addition, because the test must include the possibility of disconfirming the prediction, and thus the theory or evidence (e.g., argument of the generalizability of an effect) used to make the prediction, researchers should entertain the

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

alternative explanations for a null finding besides that the theory is incorrect. Indeed, it is difficult to disseminate a null result because it is difficult to eliminate alternative explanations for it (e.g., weak manipulation or measurement; low power). Finally, one should, during the planning stage, consider the possibility that the direction of the effect hypothesized is the reverse of the prediction. Thus, to maximize the likelihood that one will have something reasonably definitive to say, design a study such that no matter what is found, what it means is likely interpretable.

Regardless of whether one does all of the above, for HARKing to be an issue, the evidence must not have supported the prediction. Even then, if the second element was done well, accepting the null hypothesis should not motivate HARKing because one would presumably have the interesting and interpretable finding that the theory was disconfirmed. It is only if the null hypothesis was accepted and alternative explanations for a null effect likely that HARKing of the dropping-the-hypothesis kind would be considered. That is, one might remove from a manuscript or filter a manuscript from publication when the null results are not clean enough to interpret or the most likely interpretation is uninteresting (e.g., no variance on the key variable). Alternatively, HARKing of the adding-a-hypothesis kind might be considered if the null was rejected but found to be in the opposite direction of the hypothesis. In this latter case, it also needs to be the case that the original theory was misapplied in the first place, or the particular operation of an element of the method so weak (e.g., poor manipulation) that, in hindsight, it seems puerile to have ever applied the theory or to have used the design element, *and* an alternative theory is available and not puerile, or in the case of the design element, there is some scientific value to what remains of the study. In these cases, not HARKing would result in asking the reader to indulge in one's folly for a bit before getting to the findings that might be interpretable.

Using Logic and Evidence Rather Than Order to Determine Hypotheses Reported

This is the issue for me. Sometimes in my own work and sometimes in the work of others, the results from the study show that the operationalization of a manipulation or measure is so confounded (e.g., manipulation relates to nontargeted outcome) or so weak (e.g., manipulation check null or reliability low) that it is uninformative. In other cases, the folly of the original theory and hypothesis argues for its removal from the dissemination of the findings of the study. In the first case, it is simply to expunge the hypothesis because the test of it would be silly or, worse, misleading. For example, one finds very little variance on a variable, which appears to be via an unanticipated ceiling or floor effect, and thus the null hypotheses for all the tests with this vari-

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

able are accepted. I vote for pretending the whole affair never happened. In the second case, I would argue against including the poorly conceptualized original theory and hypothesis, not because it makes the author look bad but because it creates an unnecessary burden for the reader and increases the chance of misinterpretations, undermining the dissemination of the research.

Indeed, what is likely the most common form of HARKing is the revision of hypotheses, not because they were not supported, but because as stated they are not as diagnostic of the theory or research question as an alternative hypothesis that can be tested given the methods. That is, it is not the theory, its application to the study context, or some method element that is puerile; it is the information value of the test articulated in the hypothesis that is not as high as it could be. In some cases, this information undervalue is not clear until steeped in the process of analyzing the data and interpreting the results. Should that weakness in one's ability to foresee the best test force readers to suffer through those less informative tests?

Perhaps it seems that I am advocating lying to the reader regarding when the hypothesizing was done. Regarding this point, here is what I tell my students. When HARKing, saying "therefore, I hypothesized" *unnecessarily* fabricates a sequence of events. The hypothesis is *not* about the author. It is about the theory or arguments brought to bear to the research question constrained by the methods used to test the theory. These theories or arguments might be brought to bear after the results are known; but, regardless of order, the hypotheses follow from these arguments, and that is how to state it (e.g., "this leads to the following hypothesis"). Indeed, the validity of this statement is what needs to be evaluated. If the previous arguments do not lead to the hypotheses, it appears we have folly. Importantly, the ability to evaluate the folly of the revised hypotheses will be much more readily evaluated when presented in the introduction than the discussion because the focus of the introduction is about arguing for the predictions. If saved for the discussion, much less space will be devoted to vetting the arguments (of course, empirical data would be helpful here, but I will boldly predict than the space devoted will vary dramatically and the direction will be that more will be presented in introductions).

"But, professor," one of the cleverer students begins, "omission is a form of deception, is it not?" Herein lies the conundrum. Yes, omission *can* be a form of deception. Our institutional review board (IRB) documentation points this out. If one is videotaping a data collection session surreptitiously, failing to reveal that in a debriefing is unethical. However, not divulging the hypotheses in the consent form is also an omission that is rarely deception. It is a detail. If that detail relates to risk, omitting it compromises informed consent and is unethical. Omitting the order of hypothesizing is deception if

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

it is hiding a detail that can impact the interpretation of the results (e.g., the experiment-wise error). However, including a detail can sometimes undermine the benefit. In the case of revealing hypotheses to pending participants, it may increase the possibility of demand characteristics (i.e., the effects observed are biased by the context of the study). Including the order of hypothesizing, I suspect, will bias beliefs regarding the validity of the findings. That is, it is often an irrelevant cue that is used in a judgment, making the judgment biased. It would be nice if I had data to back that statement up, but I do not yet.

Instead of order of hypothesizing, one should use as cues (a) the logic of and evidence for the arguments leading up the hypothesis and (b) the rigor of the methods in terms of drawing inferences about the results. This gets us back to the advice to think through the possible results to make the methodology rigorous enough to interpret whatever is found. However, such forethought is very hard to do well. Moreover, the costs associated with all the design elements that might be employed to optimize the rigor across all the possibilities are simply too high or compromise other validities (i.e., all studies involve a set of tradeoffs; Runkel & McGrath, 1972). Thus, from time to time, we might find ourselves reviewing our analyses or a paper from another where the results took an unexpected turn. What then to do?

If the logic of and evidence for the arguments leading up to the original hypothesis were sound (i.e., not puerile), and the rigor of the methods allows reasonable interpretation of the finding (e.g., the manipulation checks confirmed the manipulations), one has an interesting story to tell in the disconfirmation of the application of the logic and evidence to the context of the study (i.e., do not HARK). If, upon hindsight—perhaps motivated by the findings, perhaps not—the logic and evidence were not sound, or upon reflection of more logic and evidence—but not the evidence from the current study—it appears that the original logic and evidence do not apply to the context studied, as in the case where a reviewer skewers your logic, then drop it (i.e., HARK). If no other reasonable logic applies or if the logic applied by the author you are reviewing is not sound *regardless of the order of the logic and data collection* (which you might not know anyway as a reviewer), do not disseminate it (i.e., HARK by not publishing it). But, if revised logic and evidence for a set of hypotheses are sound and the method allows for the elimination of most of the alternative explanations, write it up as such or advocate for the author doing so as the reviewer (i.e., HARK).

Now, because science is complex, there are a few more qualifications. If the purpose of the study is assessing a predictor's validity (e.g., for selection purposes), then knowing that it does not work or work as predicted in some context is helpful, and it should be reported (Vancouver, 2010). If

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

assessing an intervention is the purpose and the finding is that it is not effective, one should only suppress if there is strong evidence that the intervention was compromised in some way (e.g., very poor fidelity; evidence of treatment diffusion). Even then, it might be informative to report the susceptibility of the intervention to these issues. Of course, sometimes it is just poor execution, and another study showing that poor execution undermines interventions generally, as opposed to this one specifically, may not be informative.

Of course, an alternative explanation difficult to eliminate is chance. Unless the process (i.e., theory) is deterministic, all the factors known, and all the measurement valid, chance will always play a role. Capitalizing on chance, though, will only play a role when conducting multiple tests. If a hypothesis is made, but the opposite effect is found, chance might be the reason, but the probability of this is reflected in the *p*-value. It is only when multiple tests are used that *capitalization* of chance occurs. Even then, it seems to me that the chance explanation is getting a myopic piece of this discussion. I am trying to provide some perspective. In particular, a factor that strongly determines the likelihood that a finding is disseminated and taken seriously is the quality of the writing. Teaching one to tell a clean, linear story with a cogent argument leading to a set of hypotheses when writing up a study is not the bane of good science; it is part of what a good scientist needs to do. Moreover, it should allow for the proper evaluation of the results.

To understand why perspective is needed requires considering the various purposes of research. For instance, teaching about the dangers of HARKing and data cooking (i.e., filtering studies or a study's results in some way) is important when one is trying to estimate effects for practical purposes (e.g., validity coefficients or intervention effectiveness). For example, publishing only when some presumed intervention showed an effect while suppressing trials that do not show an effect is a serious issue. However, researchers are often trying to figure out how to test process theories involving latent constructs using participants who have a lot driving their behavior besides the task laid out by the researcher. Yet, when the data indicate that psychological realism or other aspects of the protocol are not lining up, it seems it would lead to a lot of useless publications if forced to present the results. Indeed, taken to its logical conclusion, no editor should reject any paper because it could mean withholding one or a set of hypotheses. Anything and everything would be published because nondiagnostic hypotheses could not be changed and weak methods could not be used as an excuse for dropping hypotheses.

In scientific writing, authors need to position their arguments within the current literature to (a) explain the importance of the topic, (b) identify the potential contribution of the scholarly work, and (c) build a robust argument

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press

for the position taken and, in empirical work, how the data collected inform the research questions. When considering these goals, I cannot make the argument that HARKing makes a substantive impact one way or the other, particularly in the grand scheme of improving the practice of science. Rather for me, it misdirects focus from what is important: the quality of thinking, theory, and methods. That is, *if* we had stronger thinking, theory, and methods, HARKing would not be a problem. Moreover, my basic argument is that *when* we have stronger thinking, theory, and methods, HARKing is not a problem. Rather, in that case, the suppression of HARKing and some of the other "questionable research practices" (Johns, Loewenstein, & Prelec, 2012) is more likely the problem.

Science is complex. When designing studies to diagnose theories, a priori hypothesizing is an important strategy. Tactically, HARKing might make sense when an unexpected finding from a methodologically strong study leads to a new insight or a more accurate understanding of where a theory applies or a methodological or logic weakness becomes known after the results are known. So what is a good scientist to do? Grand et al. (2018) provide much of the answer. Design and evaluate studies based on relevance and methodological rigor. To that I would add that one should carefully evaluate the logic of the theory or arguments used. If everyone does those three things, we will have better science.

References

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology, 69*, 709–750.

Grand, J. A., Rogelberg, S. G., Allen, T. A., Landis, R. S., Reynolds, D. H., Scott, J. C., . . . Truxillo, D. M. (2018). A systems-based approach to fostering robust science in industrial-organizational psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 11*(1), 4–42.

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management, 43*, 5–18.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524–532.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality & Social Psychology Review, 2*, 196–217.

Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to method*. New York: Holt, Rinehart, and Winston.

Vancouver, J. B. (2010). Improving I-O science through synthetic validity. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*, 360–362.

https://doi.org/10.1017/iop.2017.89 Published online by Cambridge University Press