

RUNNING HEAD: REALIZING RECOMMENDATIONS

**Forwarding the ACBS Task Force recommendations: The case for the Functional-
Cognitive framework and out-of-sample prediction**

Jan De Houwer* & Jamie Cummins*

Ghent University, Belgium

mailing address:

Jan De Houwer
Ghent University
Henri Dunantlaan 2
B-9000 Ghent
Belgium
email: Jan.DeHouwer@UGent.be
phone: 0032 9 264 64 45

Abstract

It is important that scientists reflect on their scientific aims and on how to achieve those aims. The report of the ACBS Task Force on the strategies and tactics of CBS research is an important document in that it provides explicit recommendations on what is needed to realize the aims of CBS. In this invited commentary on the report, we reflect on two ways in which several of the recommendations in the report can be fulfilled. More specifically, we specify the ways that the Functional-Cognitive framework can be used to foster communication and collaboration at different levels of analysis, and the ways in which modern data scientific approaches (specifically out-of-sample prediction) can be used in the context of existing statistical methods to provide ideographic prediction.

.

In recent years, it has been argued that psychology is in crisis. Most often, this claim is made within the context of empirical research (i.e., the fact that many of the findings reported in the psychological literature cannot be replicated; e.g., Open Science Collaboration, 2015) but similar claims have been made within the context of theory building (i.e., that psychology lacks good theories; e.g., Eronen & Bringmann, 2021) and measurement (i.e., that psychology lacks good measures; e.g., Flake & Fried, 2020). Although these crises are genuine and important, one could argue that they are trumped by a crisis of purpose. Psychological scientists are rarely explicit about what their aims are or how they can assess whether their research brings them closer to their aims. When the aims of a scientific discipline are unclear or when it is unclear how progress in realizing these aims can be assessed, it is unlikely that this scientific discipline will progress in a systematic manner. Empirical research, theory building, and measurement are all at the service of reaching scientific aims. Hence, even when the crises in empirical research, theory building, and measurement are solved, scientific progress will still depend on having clear scientific aims and indices of progress in reaching these aims.

With this in mind, it is admirable that the community of Contextual Behavioral Science (CBS) researchers is explicit about its scientific aims (Hayes et al., 2012). It is also admirable that they reflect and communicate on a regular basis about what is needed to achieve those aims. We therefore welcome and applaud the efforts of the authors who contributed to the report of the ACBS Task Force on the strategies and tactics of CBS research (Hayes et al., 2021). The report is, of course, only a first step in a much more elaborate process (see De Houwer et al., 2022, for a reflection on this process).

In this invited commentary, we focus on two aspects of the recommendations laid out in the report: namely, on fostering collaboration at different levels of analysis of psychological phenomena (Recommendations 1 and 22) and on the need for appropriate statistical methods for individual-level, ideographic analysis (Recommendations 14 and 19). In terms of moving the recommendations made in the target paper forward, we offer two insights. Firstly, for recommendations relating to the need for multilevel/multidisciplinary coordination, we would suggest that the use of functional-cognitive frameworks may directly catalyse such cooperation (De Houwer, 2018). Secondly, for recommendations relating to the need for individual-level prediction, we would suggest that (in addition to the development of new psychometrics), that CBS researchers also refine their approach to existing group-level statistics, embracing a more continual approach to model estimation and out-of-sample prediction (e.g., Tashman, 2000). In the following sections, we elaborate on these points in more depth.

Two recommendations explicitly alluded to the need for greater cross-discipline communication and the ability to coordinate between different levels of analysis (Recommendations 1 and 22). In this regard, we would advise that CBS researchers consider the use of the functional-cognitive framework (De Houwer, 2011, 2018; Hughes et al., 2016) as a lens to enable more fluent communication between different fields (and indeed, support the mutual building of knowledge for each of these fields). In essence, the functional-cognitive framework begins from a similar premise to that of the recommendations made in the target article: namely, that one can view descriptively identical phenomena from multiple levels of analysis, and that these levels of analysis can be mutually supportive (rather than mutually exclusive; De Houwer, 2011).

What the FC framework adds is contextualization over the precise ways in which the functional and the cognitive level of analysis relate to one another. It defines these levels in terms of scientific goals. Whereas functional researchers aim to predict-and-influence behavior by documenting ways in which behavior is a function of the environment (e.g., Hayes & Brownstein, 1986), cognitive researchers aim to document the mental mechanisms via which elements in the environment influence behavior (Bechtel, 2005, 2008). Provided that these levels are kept distinct, research at one level can benefit from research at the other level. More specifically, whereas knowledge about environment-behavior relations constrains cognitive theories about mental mechanisms, those theories can provide inspiration for discovering new environment-behavior relations. Moreover, because both functional and cognitive researchers collect data about environment-behavior relations, they can also learn from exploring each other's empirical research (see De Houwer, 2011, and Hughes et al., 2016, for more details).

Let us consider one example: research into stimulus relations. CBS researchers in general are interested in examining the functional environment-behavior relations which give rise to particular relational responses and relational learning histories. Many within CBS may believe CBS and behavior analysis are the only fields which deal with such phenomena; however, this is certainly not the case. As De Houwer (2018) previously highlighted, cognitively oriented researchers spanning from developmental science, comparative psychology, educational science, and analogical reasoning research have substantial interests in the examination of relational responding. We would argue that the idea that only those within CBS/behavior analysis are interested in the study of stimulus relations in large part comes from the fact that these disciplines vary in the nomenclature that they use. Whereas we in CBS may refer to a phenomenon as “analogical relational responding”, others

differentially refer to this same phenomenon as “analogical processing”, “relational concepts”, and various other terms, and typically situate their analysis at the cognitive level (i.e., they are interested in proposed mental events which mediate the environment-behavior relations involved in stimulus relations).

In addition to clarifying these alternative levels of analysis, adopting the FC framework in this regard can give, and has already given, rise to productive and novel insights for both cognitive and behavioral researchers. For cognitive researchers, for example, concepts from Relational Frame Theory (RFT; Hayes et al., 2001) have led to novel insights into phenomena such as the Stroop effect. In a recent paper, Liefvooghe and colleagues (2021) adopted a functional approach to the Stroop Task and demonstrated that Stroop effects can emerge for derived stimulus relations. This, in turn, offered substantial implications to cognitive theories relating to automaticity, which had previously considered only directly trained stimulus relations in the emergence of automatic responses.

On the other hand, CBS researchers (through the utility of the FC framework) may also benefit from findings derived from cognitive science. For instance, an increasing focus is emerging within CBS regarding the assessment of relational responding repertoires. Relational assessments such as the Relational Abilities Index (RAI; Colbert et al., 2020) have been developed in this regard. Yet much more sophisticated assessments of relational responding have already been developed within cognitive science which are arguably more amenable to a functional approach than current measures used within CBS. The RAI currently assesses 5 relational skills: opposition, difference, temporal, quantity, and analogical relational responding. However, one may note that these relational skills are delineated based on topography, rather than based on function. Indeed, an implicit

assumption that runs through much of RFT research is that topographically different relations correspond to separate functional response classes, even though this may not necessarily be the case. By contrast, the Test of Relational Reasoning (TORR; Alexander et al., 2015), a measure developed from a cognitive perspective on relational reasoning, delineates its subscales based on functionally distinct relational operant classes derived based on comparison of multiple factor analysis models (namely: analogical, anomalous, antinomial, and antithetical relational reasoning). Of course, it may not necessarily be the case that this delineation is most useful to CBS researchers. However, CBS researchers can certainly learn from this approach and the accompanying methodologies.

Indeed, in addition to these direct contributions of the FC framework, we believe that such an approach may also indirectly facilitate the uptake of other Recommendations within the report. For example, Recommendation 15 emphasizes the need to integrate experimental research findings into applied work. However, many applied practitioners outside of CBS speak the language of cognitive, rather than functional, psychology (see De Houwer et al., 2017). The use of the FC framework would allow CBS to communicate with those non-functional practitioners by both (i) helping them to identify the core phenomenon of interest, and (ii) using the language with which they are familiar and appealing to the epistemological values which they hold. This, in turn, would better enable the scalable adaption of CBS findings into applied contexts beyond the clinics of CBS practitioners alone. Indeed, this would likely also provide these non-CBS practitioners with a greater understanding and appreciate of CBS methods, by allowing those practitioners to understand the similarities and differences of functional and cognitive psychology in more concrete terms, as well as aiding in the communication of functional research findings to cognitive audiences and vice versa. Such a mutually supportive research environment is precisely the goal of the FC framework.

While the functional-cognitive framework may provide scaffolding to foster communication across different disciplines and levels of analysis, this speaks less to the methods which empirical researchers may use daily within their own work. Indeed, two other Recommendations from the report appeal to a very different side of practice: the need to develop alternative idiographic psychometrics (Recommendation 14) and to develop measures and procedures for prediction at the individual level (Recommendation 19). We wholeheartedly agree with the importance of these recommendations. However, we reject the implicit premise laden within these recommendations that extant methods are not up-to-task for these issues. While the development of new methods for these purposes is worthwhile, we would also advocate that CBS embrace extant methods and utilize them in ways that they have not been to-date.

In concrete terms, what we are saying is that existing statistical methods *can* be used for making individual-level predictions, but we simply have not used them in this manner to date. Let us illustrate this point by firstly outlining the modal use of statistical modelling within CBS (and indeed, psychological science more generally), and then compare this to how we might otherwise use modelling. Suppose a researcher is interested in examining whether scores on a measure of relational responding predict performance in educational outcomes (e.g., as in Colbert et al., 2020). In general, the lifecycle of creating and using a statistical model for such a research question might be summarized as follows:

1. Collect some data,
2. Fit a statistical model (e.g., univariate linear regression) to these data,
3. Make theoretical/conceptual interpretations of p-values associated with parameters in that model (i.e., examining if relational responding predicted educational performance),

4. Write-up these results in a paper.

This process, we agree, is wholly insufficient for learning about idiographic processes.

However, we would argue that this insufficiency is not due to a limitation of the statistical methods themselves, but rather a failure to effectively utilize them to their fullest extent. A meteorologist, for example, can make highly specific and accurate predictions (e.g., what the weather will be like in Dublin City Centre next Tuesday between 5pm and 6pm) using these same statistical techniques. Likewise, social media platforms such as Twitter, Facebook, and Instagram are notorious for their highly targeted advertisement, which they also achieve using these same statistical methods. Indeed, statistical models are used in countless other contexts to make highly specific, fine-grain predictions about individual units of analysis (Taylor & Letham, 2018). In fact, a small minority of researchers active within the CBS community have previously used such methods (e.g., Whelan & Garavan, 2014; Lespine et al., 2022). Nevertheless, the use of statistical modelling in this manner is relatively underused within CBS.

CBS practitioners can more effectively utilize statistical models for individual-level prediction by reconceptualizing a statistical model as *a tool for prediction* rather than as a description of a relationship between variables. In this case, fitting a statistical model is equivalent to sharpening the tool; it merely represents the first step in using the tool. After fitting the statistical model, the researcher can then *use* the statistical model to make predictions on new data. One common approach to such a perspective within data science is the use of a “train-test split” or “holdout dataset” (Langford & Schapire, 2005). To return to the earlier example of relational responding and educational performance, this approach would involve the following steps:

1. Conduct an *a priori* power analysis to determine the sample size required to sufficiently power the planned analysis for a smallest effect size of interest,
2. Collect some data which meet these criteria for power,
3. Split these data into a training dataset (80% of the sample) and a testing dataset (20% of the sample),
4. Fit a statistical model to the training dataset,
5. Make theoretical/conceptual interpretations of p-values associated with parameters in that model (i.e., examining if relational responding predicted educational outcomes),
6. Identify how well this model predicts outcomes *for individuals* within the testing dataset using relevant evaluation metrics,
7. Write these results up as a paper,
8. Make the final model openly available.

Although these steps are similar to the modal workflow in CBS, they come with some critical additions. The specification of an *a priori* power analysis is used to highlight the minimum sample size required for a specific test of interest (note that this power analysis should dictate the size of the *training* dataset). While some may argue that it can be difficult to know what effect size one might expect, opting for a smallest effect size of interest addresses this issue (see Lakens et al., 2020, for an elaborated discussion of the SESOI). Additionally, the split of the data into separate training and testing datasets allows for a clear delineation of which data are to be used to teach the model, and which data are to be used to evaluate the predictions of that model.

Theoretical interpretations regarding parameter values can still be made. However, the model is also evaluated in terms of its “out-of-sample prediction”; i.e., how well the model

predicts new observations. If the model predicts these new observations well, then the model is successfully achieving individual-level prediction. If the model does not predict these new observations well, then this suggests that the model may be *overfit* (i.e., it has been conditioned too strongly on the training data and consequently it is not generalizable to observations outside of the training data; Cook & Ranstam, 2016). In this way, we can evaluate how well the model performs at individual-level prediction and adjust in cases where the model is underperforming. It is important to note that the performance of this model in terms of prediction is not evaluated through null hypothesis significance testing metrics such as a p-value or effect size; rather, the predictive accuracy of the model can be evaluated using metrics such as the mean absolute error of predictions or the mean squared error of predictions (for linear regression models), the precision/accuracy predictions (for logistic regression models), or several of a multitude of other metrics (for a review of potential metrics, see Botchkarev, 2018). Which method is most appropriate, and what values constitute “good” performance, will necessarily depend on a myriad of factors, such as the researcher’s resources and goals. However, the above outlined approach is used successfully by many other fields to make the kinds of individual-level predictions toward which CBS strives, and so too can CBS.

Importantly, we can also make the final model openly available to other researchers, statisticians, and practitioners. In this way, others may improve our existing model and/or use the model to make individual-level predictions in their own studies. Such an iterative process to model development and deployment is common in the world of data science, and CBS would do well to replicate such approaches. CBS need not reinvent all of statistics to achieve the individual-level predictions which other fields are already achieving; rather, it needs

simply to recognize that the modal way statistics are used within the field is not fully realizing the potential of those statistics.

To sum up, we agree wholly with the recommendations made within the Task Force Report. However, we also believe that CBS as a community need not (re-)invent new ways of communicating or statistical methods to achieve many of these recommendations. To foster collaboration and understanding of phenomena at different levels of analysis, one may look to the FC framework to provide verbal and conceptual scaffolding which can be used to navigate between these levels in a collaborative, constructive manner. Likewise, for ideographic measurement and prediction, we need not develop a new class of statistics, nor learn about entirely new branches of statistics. Rather, we need only re-conceptualize the ways in which we approach the current statistical methods which are already common in CBS and emphasize the need for a greater focus on train-test splits and the use of out-of-sample prediction. In embracing these existing ways of thinking, CBS will already go a long towards achieving several of the recommendations from the Task Force report.

References

ACBS (2021).

https://contextualscience.org/news/adoption_of_open_science_recommendations

Bechtel, W. (2005). The challenge of characterizing operations in the mechanisms underlying behavior. *Journal of The Experimental Analysis of Behavior*, 84, 313–325.

Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Routledge.

Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. arXiv preprint

arXiv:1809.03006.

Chester, D. S., & Lasko, E. N. (2021). Construct validation of experimental manipulations in social psychology: Current practices and recommendations for the future.

Perspectives on Psychological Science, 16(2), 377–395.

<https://doi.org/10.1177/1745691620950684>

Cook, J. A., & Ranstam, J. (2016). Overfitting. *Journal of British Surgery*, 103(13), 1814–1814.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, 6, 202–209.

De Houwer, J. (2018). A functional-cognitive framework for cooperation between functional and cognitive researchers in the context of stimulus relations research. *The Behavior Analyst*, 41, 229–240.

De Houwer, J., Cummins, J., Boddez, Y., & Finn, M. (2022). The uptake of recommendations requires not only clarity on what is recommended but also on why it is recommended, how it can be realized, and who needs to take action. *Unpublished manuscript*.

De Houwer, J., Hughes, S., & Barnes-Holmes, D. (2017). Psychological engineering: A functional-cognitive perspective on applied psychology. *Journal of Applied Research in Memory and Cognition*, 6, 1–13.

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, 16(4), 779–788.

<https://doi.org/10.1177/1745691620970586>

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement

- practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3, 456–465. <https://doi.org/10.1177/2515245920952393>
- Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *The Behavior Analyst*, 9, 175-190.
- Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science*, 1, 1–16. <https://doi.org/10.1016/j.jcbs.2012.09.004>
- Haynes, S. N., Mumma, G. H., & Pinson, C. (2009). Idiographic assessment: conceptual and psychometric foundations of individualized behavioral assessment. *Clinical Psychology Review*, 29(2), 179–191. <https://doi.org/10.1016/j.cpr.2008.12.003>
- Hughes, S., De Houwer, J., & Perugini, M. (2016). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*, 51, 4-14.
- Langford, J., & Schapire, R. (2005). Tutorial on Practical Prediction Theory for Classification. *Journal of Machine Learning Research*, 6(3), 273-306.
- Lespine, L. F., Rueda-Delgado, L., Vahey, N., Kiiski, H., Enz, N., Boyle, R., ... & Whelan, R. (2022). Changes in inhibition-related brain function and psychological flexibility during smoking abstinence predict longer time to relapse. Preprint. <https://psyarxiv.com/472p5/>
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437-450.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45.
- Whelan, R., & Garavan, H. (2014). When optimism hurts: inflated predictions in psychiatric

neuroimaging. *Biological Psychiatry*, 75(9), 746-748.

Acknowledgements

Jan De Houwer, Jamie Cummins, Ghent University, Ghent, Belgium. Jan De Houwer and Jamie Cummins contributed equally and agree to shared first authorship. Correspondence should be addressed to Jan De Houwer, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. Electronic mail can be sent to Jan.DeHouwer@UGent.be . The preparation of this paper was made possible by Ghent University Grant BOF16/MET_V/002 to Jan De Houwer. We thank the other members of the Learning and Implicit Processes Laboratory at Ghent University (www.liplab.be) for their input during a discussion of the report of the ACBS Task Force at one of our lab meetings.

Funding Information

The preparation of this paper was made possible by Ghent University Grant BOF16/MET_V/002 to Jan De Houwer.

Declaration of interest: none