European Journal of Personality, *Eur. J. Pers.* **34**: 593–598 (2020)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.2310

Editorial

A (More) Behavioural Science of Personality in the Age of Multi-Modal Sensing, Big Data, Machine Learning, and Artificial Intelligence

Over more than a decade ago, Baumeister, Vohs, and Funder (2007) asked: whatever happened to *actual behaviour* in psychology? Shortly after, Furr (2009a) published a target article here in *EJP* on personality psychology as a truly behavioural science and discussed the (low) prevalence, meaning, importance, and measurement of behavioural assessment (see also Furr, 2009b). Now, it is time to take stock: have we, as personality psychologists, moved towards using more behavioural assessment and understanding what people are *actually* doing in their personal, social, and occupational lives? Has new technology helped us achieve these goals?

From the available literature, it would seem that behavioural measurement has been slowly on the rise but also that we still only rarely use actual behaviour as a data source (Nave, Feeney, & Furr, 2018). However, rapid technological advancements, the digitization of daily life, the ubiquity of data-gathering tools (e.g. smartphones), and the introduction of multi-modal sensing methodology, 'big data' applications, machine learning, and artificial intelligence approaches to psychological science promise more and perhaps even better behavioural assessment than was possible decades or even just some years ago. For example, what people speak (e.g. from audio snippets), write (e.g. on online social networks, text-massaging apps, and emails), or do (e.g. captured in videos via minicameras, geospatial movements via GPS, app usage on the smartphone and internet, gameplay, or pictures of activities posted on online social media platforms) in their daily lives yields intensive and massive amounts of data that need to be mined and modelled appropriately (Blake, Lee, de la Rosa, & Sherman, in press; Harari et al., 2016; Ilmini & Fernando, 2017; Kosinski, Stillwell, & Graepel, 2013; Onnela & Rauch, 2016; Vinciarelli & Mohammadi, 2014). Thus, to some, it may seem that personality research and assessment in particular, but probably basic and applied psychological science in general, may be entering a new age where large amounts of data on people's actual behaviours, or traces of their behaviours, can be sampled or mined and modelled with sophisticated algorithms (Chamorro-Premuzic, Akhtar, Winsborough, & Sherman, 2017; Kosinski, Wang, Lakkaraju, & Leskovec, 2016; Mahmoodi, Leckelt, van Zalk, Geukes, & Back, 2017; Montag & Elhai, 2019; Tracey, 2020; Woo, Tay, & Proctor, 2020; Wright, 2014).

OVERVIEW OF THE SPECIAL ISSUE

The current special issue is dedicated to exploring a more behavioural science of personality in the age of multi-modal sensing methodology, big data, machine learning, and artificial intelligence: how is sensed and/or big data 'behavioural', which behaviours are captured, how can such behavioural information be analysed and used, and how does all of this inform (or transform parts of) personality psychology? In total, this special issue consists of 19 stellar papers which fall broadly into theoretical/conceptual/review and empirical papers. Five review-type papers deal with conceptual, theoretical, and 'big picture' issues (conceptualizing personality via big behavioural data: Boyd et al.; personality research and assessment using big data, sensing methods, and machine learning: Stachl et al.; Alexander et al.; Harari et al.; and idiography and renewed person-centred focus: Renner et al.). These papers not only give good and highly accessible overviews of past and recent research or methodologies and thus set the stage for, and contextualize, the subsequent empirical articles; they also provide a wider picture by discussing various merits, limitations, pitfalls, and implications (e.g. ethical, legal, social/societal, and practical). Each of these five papers has a different focus and provides different key insights, yet the papers also overlap in certain aspects (e.g. the repeated plea that psychometrics and assessment need to be taken seriously). The latter fact is good because it shows that different teams of researchers operate from a common understanding and share common standards.

Fourteen papers are empirical in the sense that they report analyses of data. However, many of these papers also make cogent conceptual and theoretical observations beyond just the data they have at hand. They thus nicely complement the other five papers that did not use data. The empirical papers can be broadly classified into (i) dealing with behavioural sensing (in the broadest sense) or (ii) using digital footprints and text-based analyses.[1] The behavioural sensing articles present cutting-edge research using the following different methodologies and data sources: smartphone sensing (extracting personality state data: Rüegger et al.; relations between mobility behaviours and subjective well-being: Müller et al.; and individual differences and day–night behaviours: Schödel et al.), the Electronically Activated Recorder (relations between traits, daily behaviour, and language use:

[1] This distinction is somewhat arbitrary, but still serviceable for the sakes of structuring the empirical papers in this special issue.

Mehl et al.; and patterns of daily socializing: Danvers et al.), physiological signals (skin conductance in arousing events: van Halem et al.), and performance in behavioural tasks (conscientiousness-related response behaviours: Gniewosz et al.). The digital footprint and text-based articles present cutting-edge research using social media, digital footprints, and online behaviours (social media text mining for personality assessment: Tay et al.; fluctuations in states using social media language: Weidman et al.; prediction of online hostility: Rosenbusch et al.; and prediction of trait nuances from digital footprint data: Hall et al.) as well as harnessing text-based approaches outside of (or at least not solely focused on) social media (deriving a personal values dictionary from different texts: Ponizoviskiy et al.; extracting information on prosociality from qualitative data: McAuliffe et al.; and extracting personality-relevant information in vocabulary used in novels: Fischer et al.). As can be seen from this impressive line-up, the empirical articles span a wide range of topics and methodologies and thus demonstrate how rich behavioural and digital assessment methodologies can be.

SOME OBSERVATIONS AND RECOMMENDATIONS

Many papers in this special issue of course also highlight and tackle problematic issues or limitations of behavioural sensing, big data, and algorithm-based approaches, but their common tenor is still one of optimism and enthusiasm (which I also share). Indeed, there is much to be excited about being a personality scientist in these times, and many new challenges await (e.g. measuring states as actual behaviours and extracting trait-relevant information from these; providing more formalized theoretical accounts of the processes that link psychological experiences and overt manifestations in behaviour; paying attention to psychometric properties in big data applications; and applying all of this knowledge to solve real-world problems). As such, I felt incredibly fortunate to have been able to edit and accompany the papers that were submitted to this special issue. Reading all of these papers was extremely stimulating and also educating. I continue to be amazed what has become possible and very much look forward to what we will achieve in the future, and I hope readers will feel the same way after reading the papers in this special issue.

My enthusiasm notwithstanding, I would like to share some broader, and perhaps more critical, observations that can be of course also found discussed or addressed in the following papers (with much more sophistication and nuance than I could summarily present here). Readers may keep these in mind when going through the special issue. In my eyes, the papers have done a great job in circumventing usual 'traps' (see some below) and highlighting the true potential that behavioural sensing, big data, and algorithm-based approaches harbour for personality science. As such, the following five observations, which also serve as recommendations for moving forward, may set the stage for reading, and especially appreciating, those papers.

Observation #1: focus on psychometrics in general and nomological validity in particular

First, it is important to also apply stringent standards of rigorous psychometric testing (e.g. estimating different forms of reliability and validity) to algorithm-based trait or state scores extracted from sensed behavioural and big data (Bleidorn & Hopwood, 2019). For example, in the past, less attention was devoted to also showing discriminant validity when examining convergences between self-reported trait scores and algorithm-based trait scores (e.g. extracted from digital footprints). While certain forms of reliability may be easier to achieve because of ever more precise measurements (e.g. via sensing signals and wearables), especially validity is a key concern. Are we really measuring what we want to measure or purport to measure? Do our 'new' behavioural measures also capture variance from our 'old' or gold-standard measures, which are often self-reports? And should they do that in the first place, or should we best consider them different proxies of the same interesting phenomena? What are their predictive abilities? And what can they be used for? It is especially pressing to demonstrate what exactly we are measuring (construct validity) and what those measurements can be used for (criterion validity and utility). Answering these and many more psychometrics-related questions (e.g. Ziegler, 2014) will be important to move forward if algorithm-based scores should be used for proper personality assessment. Several of the papers included in this special issue also make this case and show how we can approach such questions.

Associated with the previous remarks is that there needs to be a stronger focus on the nomological validity of algorithm-based trait and state scores. Understandably, many studies have so far focused almost exclusively on convergent validity (most often the association between a self-reported and algorithm-based trait score; see also the second observation discussed below). However, as the papers in this special issue nicely illustrate, there are so many more things one can do with sensed or algorithm-based scores. Thinking in terms of nomological networks (Cronbach & Meehl, 1955), it will be important to piece together the causes (or antecedents), concomitants (or correlates), and consequences (or outcomes) of sensed or algorithm-based scores. This will not only serve to elucidate better what we are measuring and reduce jingle fallacies (i.e. two different phenomena obtain the same label) and jangle fallacies (i.e. identical phenomena obtain different labels) but also contextualize those scores better in terms of what they capture and what not (see the second observation below) and perhaps even move us towards more theory-building (see the fourth observation below).

Observation #2: clarity on what we are measuring or approximating and what to do with those measurements

Second, as already stated in and closely tied to the preceding points, several studies so far have been concerned with convergent validity in the sense that algorithm-based trait scores need to approximate some criterion, which has mostly been

DOI: 10.1002/per

self-reported trait scores. This is no doubt a first and important step which could be considered the 'proof-of-concept' phase. An important question, however, is whether we at all should want to re-tap self-reports. For example, why go through the trouble of getting all that difficult-to-sample and perhaps ethically complicated data (from mobile sensing, digital footprints, wearables, etc.) when—eventually—we will have an algorithm try to approximate the self-report as best as it can? It would have been easier to just sample that self-report (if it is really what we are interested in). Additionally, current algorithms do seem to outperform most kinds of human raters in predicting some self-reported trait scores (e.g. Youyou, Kosinski, & Stillwell, 2015), but they usually require a lot of data to do so (e.g. several hundreds of Facebook Likes), are measured against the better human judges (e.g. spouses), and show predictions (which can be interpreted as convergent validity) that do not lend us to believe that really 'the same thing' is tapped that the self-report taps. On top of that, the self-report is reified as the gold standard because variance from algorithm-based scores needs to overlap as much as possible with variance from self-report scores (although there are computer science studies using other-reports as criteria also; e.g. Pianesi, 2013; Vinciarellei & Mohammadi, 2014). Lastly, by trying to optimize the prediction of self-reports, construct validity in such a sense (i.e. the algorithm-based score taps the self-report) may lead to redundancy of the sampled information. Moving forward, studies employing algorithm-based traits should seek to sample personality traits not just via self-reports but also via other data sources (e.g. indirect measures, behavioural observation, neurophysiological and biochemical indicators, strangers' impressions, informant knowledge, and experience sample). Then algorithms can be trained to predict trait scores that come from multiple data sources. We can also concurrently associate different 'indicators' of the same trait with each other and try to extract common variance to tap the 'heart of the trait' (whatever that is) beyond unique measurement/method variance and unsystematic error variance. Such an approach recognizes that a construct, like a specific trait, is tied to different data sources and measurements and tries to peal it out.[2] To truly measure a trait, such data would need to be sampled also repeatedly and trait-like components disentangled from state-like components. Using all kinds of data sources and sampling these repeatedly could be very costly, but a minimum approximation would be to use at least two data sources and at least two measurement points. Here, algorithm-based scores (e.g. extracted from wearables, mobile phones, digital footprints, etc.) still offer a unique window into measuring certain behavioural aspects of a trait, but they cannot be equated as the sole measure of the trait and would be at their strongest if paired with trait-relevant data from other data sources.

The previous line of thinking was concerned with what measurements seek to approximate. Another important question is what will be subsequently done with those measurements. As the papers in this special issue nicely demonstrate, there are more things to do with algorithm-based scores than simply try to optimize their tapping of self-reports. For example, if we have a priori defined (perhaps even on theoretical grounds) which sensed signals, digital footprints, etc. are relevant manifestations of a trait, then we could try to predict these from prior trait scores (ideally measured by different data sources). Now we are using algorithm-based scores as outcomes. Conversely, we could also examine the predictive power of algorithm-based scores. For example, we could establish their stability and then have the stable portion of the score (e.g. within a latent state–trait structural equation model) predict later outcomes (e.g. well-being and mobility). Here, we use the algorithm-based scores as predictors. Using algorithm-based scores in such ways is associated with certain goals (e.g. explanation and prediction), which is touched upon in the next observation.

Observation #3: disentangling description, explanation, and prediction

Third, description, explanation, and prediction—three key goals of psychology in general—are often confused with each other or not properly distinguished. Indeed, it may be questionable if we can distinguish these three properly at all in psychology, especially because they may blend into each other and also stand in the service of each other (e.g. better description can foster better explanations or predictions: Seeboth & Mõttus, 2018). For example, one researcher's explanation may be another researcher's description (e.g. Yarkoni, 2020), and the continuation of treating ever smaller, narrower, or more molecular explanantia (the explaining phenomena) as explananda (the to-be-explained phenomena) will result in a 'turtles all the way down' scenario. Further, some would say that a successful causal explanation (e.g. B exists because of A) often implies that one would be able to form good predictions (e.g. if A causes B, then A will predict B). However, explanation and prediction—while philosophically compatible—are still distinct goals, with different methodologies and priorities (Yarkoni & Westfall, 2017). These issues notwithstanding, any given research should be at least clear in what it seeks to achieve and whether that conforms, in their understanding, more to description, explanation, or prediction.

Parts of the confusion may be inherent to some forms of (personality) psychology (e.g. Bredo, 2006), but they become even more pressing when using sensed behavioural and big data approaches. Description is actually quite fundamental and non-trivial (Gerring, 2012) as it is necessary to get a hold of the 'landscape' (i.e. to better navigate something, it does not hurt to know what all is there). With all the novel technologies, 'new' data sources, and the ease with which intensive longitudinal data can be gathered today, it is tempting to stop at the level of description and perhaps simply explore patterns in one's data. Indeed, that in itself can be interesting and revealing (e.g. Kern et al., 2014), perhaps especially when applied to single cases with idiographic

[2]There are different philosophical issues associated with this (see Yarkoni, 2020), for example, whether we think the construct actually exists and is step by step uncovered by our measurements (e.g. a realist position) or whether the construct is formed by our measurements for pragmatic purposes (e.g. an instrumentalist position).

DOI: 10.1002/per

analyses. If made transparent and adhering to constraints on generality in one's interpretations, there is nothing wrong with this, and such research should be valued as much as causal or predictive analyses. However, as many papers of this special issue illustrate, we probably should not stop there either. A next step is to understand what these patterns are, what brought them about, how they 'work' in different contexts, and which consequences they have. In other words, we will want to know more about them and *use* them. This is where description stops and utilizing (e.g. for explanation or prediction)[3] begins. If we do this and grow intricate nomological networks, we run less risk at deriving a blind neo-behaviourism that simply takes any behaviours assessed with novel technologies or big data approaches at face value and is not interested in any psychological processes, dynamics, mechanisms, or functioning.

Observation #4: the roles of conceptual precision, formalization, and theory

Fourth, it is now possible and easier than ever thanks to novel technological advances (e.g. data mining of websites and multi-modal sensing) to sample a plethora of variables from many people and run some form of supervised or unsupervised machine learning over that big data to solve certain problems—often better or more accurately than we could do with simpler or piecemeal manual analyses (e.g. when creating new short forms of scales; Dörendahl & Greiff, 2020; Jankowsky, Olaru, & Schroeders, 2020; Olaru, Schroeders, Hartung, & Wilhelm, 2019). Harnessing these new opportunities should be encouraged; in fact, they probably should even be integrated into the common methodological, statistical, and quantitative training of psychologists. They will likely become the new normal, so we should know about their relative strengths and weaknesses, how to harness them better (Jacobucci & Grimm, 2020), and when to use them and when not.[4] At the same time, their use should not succumb to a mere ritual where they are just used for the sake of themselves but nobody knows why or for what. A 'Let's see what machine learning has to say about Phenomenon or Problem X' approach *could* be a useful and informative *first* step, but it should not be the only or last step. However, I concede that there may be at least two types of psychological scientists with different views on this. Some of those interested in more basic research may object to such a mentality where simply using machine learning adds something to the literature. For them, machine learning is at best a tool to test or derive hypotheses, and the proof in the pudding lies in whether the findings (not the method) will advance theory-building and cumulative insights gleaned. In contrast, some of those interested in more applied issues may see value in big data and machine learning approaches because they could be used to pragmatically address certain questions or solve certain problems. For them, machine learning is simply useful because it can already provide solutions and show what 'works' (and what not)—and that irrespective of hypothesis testing or building towards theory. Nonetheless, both types of scientists would (hopefully) agree that we need proper conceptual precision, logical reasoning, valid measures, and good data quality to be at all able to properly 'test' any hypotheses (see Scheel, Tiokhin, Isager, & Lakens, 2020) or solve any problem, respectively. There are recent calls for more formalization and theory-building in psychology (e.g. Fried, 2020; Robinaugh, Haslbeck, Ryan, Fried, & Waldorp, 2020), and these seem to underscore that we may want to focus less on *that* something works but start understanding *why* it works. In behavioural sensing, big data, and algorithm-based approaches to personality research and assessment, we also must ask ourselves: should we focus just on what works or should we focus more on theoretical issues? I am hesitant to give a definitive answer here because I think we need to balance these strivings and have them work together, as nicely exemplified in the papers of this special issue.

Observation #5: multidisciplinarity, collaboration, and focusing on 'psychological' aspects

Fifth, setting up multi-modal sensing or big data studies requires expertise in different areas (e.g. hardware, software, ethics, law, psychological assessment, and statistics) and collaboration. It should be no surprise that the future of psychological assessment via novel technologies will have to be inherently more multidisciplinary, and this also applies to personality science utilizing such methodology. Such multidisciplinarity, where different disciplines (e.g. personality psychology and computer science) come together and study the same phenomenon or effects of interest each under their unique perspective, is a good start and already being realized right now. A next step will be to transform such work into interdisciplinary endeavours where the varying perspectives, approaches, and methodologies of individual disciplines are integrated and afford a more seamless collaboration (Stember, 1991).[5]

The importance of multidisciplinary and interdisciplinary research and collaboration notwithstanding, it is at the same time vital that we do not lose sight of being interested in, or making inferences for, *psychological* variables. While it is certainly interesting to engage into behavioural data science, we still need to bring all the methodology, analyses, etc. back 'home' to a psychological level. For example, we can sample all sorts of smartphone behaviours unobtrusively, continuously, precisely, and with high granularity (e.g. which apps have been used, what has been clicked, and how long phone calls took; Stachl et al., 2020). However, such data

[3]Another central goal of some psychological disciplines is modification (besides description, explanation, and prediction). This goal is perhaps most tied to 'utilizing' insights and knowledge. For example, artificial intelligence-optimized personal therapy plans could be used to aid systematic trait change in healthy or clinical populations.

[4]Not to unduly simplify things, but many machine learning approaches are regression based—and psychologists are already well acquainted with regression analyses. Further, some statistical techniques already in our repertoire (e.g. cluster analysis) are essentially machine learning approaches.

[5]According to Stember (1991), going even further beyond is transdisciplinarity where disciplinary boundaries are completely dissolved and a new, integrated unity is formed. This is a lofty goal, of course, but certain branches of artificial intelligence research are already approaching a transdisciplinary status.

DOI: 10.1002/per

are still psychologically speaking at the surface because they are not per se imbued with any psychological meaning (e.g. you could be talking 30 minutes on the phone with a classmate, a romantic date, or your boss, and the talk could be about any amount of things, led in more pleasant or unpleasant ways, and serve different agendas). The psychological meaning of sensed or big data variables still needs to be a focal concern for psychologists. In the same way how (I think) we in fact ought to be amazed that molecular genetic information (e.g. single nucleotide polymorphisms) is associated with molar self-reported measures of explicit self-concepts at all (e.g. Montag, Ebstein, Jawinski, & Markett, 2020),[6] we should appreciate the already many and meaningful associations between more sensed or digital variables—behavioural 'snips'—and self-reports. Are we really expecting almost perfect overlaps (even if we sample hundreds or thousands of behavioural snips)? I do not think we do, and I do not think we should. The associations and predictive abilities (in terms of variance accounted for and effect sizes) found in some of the papers of this special issue should be evaluated with this mindset. Going further, several papers do 'dig deeper' and explicitly focus on the psychological meanings of the sampled information. As such, the papers in this special issue set a good example for future studies.

CONCLUSION

If you could not already tell, I am pretty excited about this special issue.[7] The papers compiled here are cutting-edge and provide a glimpse into future possibilities for our field when we want to employ multi-modal behavioural sensing, big data, machine learning, or artificial intelligence approaches. At the same time, they also contain several caveats and warnings, and they alert us how to best leverage the many opportunities but also navigate the many challenges we may face. Algorithm-based approaches to personality research and assessment, in the broadest sense, will quite likely continue to be on the rise and garner even more interest and applications in the following years. On the one hand, this will make our field (hopefully) a bit more behavioural as we gain access to methodologies that can capture many different kinds of verbal, paraverbal, nonverbal, and extraverbal behaviours and their traces in the lab and in the field—and that with a high level of precision, granularity, and temporal resolution. On the other hand, we should not forget about basic psychometric principles; be attentive to what we are measuring and what those measurements will be used for; distinguish between description, explanation, and prediction as goals in research endeavours; strive for conceptual precision and perhaps work towards more formalization and even theories; and engage into multidisciplinary collaborations but also do not lose sight of addressing psychological questions and problems.

I believe the papers in this special issue incorporate these observations or recommendations in various forms and thus already demonstrate how a sound, rigorous, and useful behavioural personality science in the age of big data and artificial intelligence can look like. Thus, I can only hope that readers will find the papers as stimulating as I did and that the papers will inspire others to further theory-building, research, and applications.

John F. Rauthmann
Department of Psychology, Bielefeld University, Germany
Email: jfrauthmann@gmail.com

[6]It has been repeatedly bemoaned that not enough variance is explained.
[7]And I am not just saying this because I edited the papers.

REFERENCES

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Blake, A. B., Lee, D., de la Rosa, R., & Sherman, R. A. (in press). Wearable cameras, machine vision, and big data analytics: Insights into people and the places they go. In S. E. Woo, L. Tay, & R. Proctor (Eds.), *Big data methods for psychological research: New horizons and challenges*. APA Books.

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*, 190–203. https://doi.org/10.1177/1088868318772990

Bredo, E. (2006). Conceptual confusion and educational psychology. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences*, *18*, 13–16. https://doi.org/10.1016/j.cobeha.2017.04.007

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Dörendahl, J., & Greiff, S. (2020). Are the machines taking over? Benefits and challenges of using algorithms in (short) scale construction. Editorial. *European Journal of Psychological Assessment*.

Fried, E. I. (2020, February 7). *Lack of theory building and testing impedes progress in the factor and network literature*. https://doi.org/10.31234/osf.io/zg84s. In press at *Psychological Inquiry*.

Furr, R. M. (2009a). Personality psychology as a truly behavioral science. *European Journal of Personality*, *23*, 369–401. https://doi.org/10.1002/per.724

Furr, R. M. (2009b). The study of behavior in personality psychology: Meaning, importance, and measurement. *European Journal of Personality*, *23*, 437–453. https://doi.org/10.1002/per.726

Gerring, J. (2012). Mere description. *British Journal of Political cience*, *42*, 721–746. https://doi.org/10.1017/S0007123412000130

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*, 838–854. https://doi.org/10.1177/1745691616650285

Ilmini, W. M. K. S., & Fernando, T. G. I. (2017). Computational personality traits assessment: A review. *In 2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, 1–6 IEEE.

DOI: 10.1002/per

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*, 809–816. https://doi.org/10.1177/1745691620902467

Jankowsky, K., Olaru, G., & Schroeders, U. (2020). Compiling measurement invariant short scales in cross-cultural personality assessment using ant colony optimization. *European Journal of Personality*, *34*, 470–485. https://doi.org/10.1002/per.2260

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., Kosinski, M., … Seligman, M. E. (2014). The online social self: An open vocabulary approach to personality. *Assessment*, *21*, 158–169. https://doi.org/10.1177/1073191113514104

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*, 5802–5805.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, *21*, 493–506. https://doi.org/10.1037/met0000105

Mahmoodi, J., Leckelt, M., van Zalk, M. W. H., Geukes, K., & Back, M. D. (2017). Big Data approaches in social and behavioral science: Four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences*, *18*, 57–62. https://doi.org/10.1016/j.cobeha.2017.07.001

Montag, C., Ebstein, R. P., Jawinski, P., & Markett, S. (2020). Molecular genetics in psychology and personality neuroscience: On candidate genes, genome wide scans, and new research strategies. *Neuroscience and Biobehavioral Reviews*, *118*, 163–174. https://doi.org/10.1016/j.neubiorev.2020.06.020

Montag, C., & Elhai, J. D. (2019). A new agenda for personality psychology in the digital age? *Personality and Individual Differences*, *147*, 128–134. https://doi.org/10.1016/j.paid.2019.03.045

Nave, C. S., Feeney, M. G., & Furr, R. M. (2018). Behavioral observation in the study of personality and individual differences. In T. Shackelford, & V. Ziegler-Hill (Eds.), *The SAGE handbook of personality and individual differences* (pp. 317–340), *1*. Thousand Oaks, CA: Sage Publications.

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality*, *33*, 400–419.

Onnela, J. P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, *41*, 1691–1696. https://doi.org/10.1038/npp.2016.7

Pianesi, F. (2013). Searching for personality. *IEEE Signal Processing Magazine*, *30*, 146–158. https://doi.org/10.1109/MSP.2012.2219671

Robinaugh, D., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. (2020, March 25). *Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction*. https://doi.org/10.31234/osf.io/ugz7y

Scheel, A.M., Tiokhin L., Isager, P.M., & Lakens, D. (2020, in press). *Why hypothesis testers should spend less time testing hypotheses*. *Perspectives on Psychological Science*.

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, *32*, 186–201.

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., … Hussmann, H. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, *117*, 17680–17687.

Stember, M. (1991). Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal*, *28*, 1–14. https://doi.org/10.1016/0362-3319(91)90040-B

Tracey, T. J. G. (2020). We can do that? Technological advances in interest assessment. *Journal of Career Assessment*, *28*, 3–13. https://doi.org/10.1177/1069072719879910

Vinciarelli, A., & Mohammadi, G. (2014). A Survey of Personality Computing. *IEEE Transactions on Affective Computing*, *5*, 273–291.

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, *5*, 273–291. https://doi.org/10.1109/TAFFC.2014.2330816

Woo, S. E., Tay, L., & Proctor, R. W. (Eds) (2020). *Big data in psychological research*. American Psychological Association 10.1037/0000193-000.

Wright, A. G. C. (2014). Current directions in personality science and the potential for advances through computing. *IEEE Transactions on Affective Computing*, *5*, 292–296. https://doi.org/10.1109/TAFFC.2014.2332331

Yarkoni, T. (2020, September 24). *Implicit realism impedes progress in psychology: Comment on Fried (2020)*. https://doi.org/10.31234/osf.io/xj5uq

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122. https://doi.org/10.1177/1745691617693393

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*, 1036–1040.

Ziegler, M. (2014). Stop and state your intentions! Let's not forget the ABC of test construction. Editorial. *European Journal of Psychological Assessment*, *30*, 239–242.

DOI: 10.1002/per