Discovering the unknown unknowns of research cartography with high-throughput natural description

Tanay Katiyar[a], Jean-François Bonnefon[b], Samuel A. Mehr[c,d]* and Manvir Singh[e]

[a]Institut Jean Nicod, Département d'études cognitives, École normale supérieure (ENS-PSL), Paris, France; [b]Toulouse School of Economics, Centre National de la Recherche Scientifique (TSM-R), Toulouse, France; [c]School of Psychology, University of Auckland, Auckland, New Zealand; [d]Yale Child Study Center, Yale University, New Haven, CT, USA and [e]Department of Anthropology, University of California-Davis, Davis, CA, USA
tanay.katiyar20@gmail.com
jean-francois.bonnefon@tse-fr.eu; https://jfbonnefon.github.io
sam@yale.edu; https://mehr.nz/
manvir.manvir@gmail.com; https://manvir.org

*Corresponding author.

doi:10.1017/S0140525X23002170, e50

Abstract

To succeed, we posit that research cartography will require high-throughput natural description to identify unknown unknowns in a particular design space. High-throughput natural description, the systematic collection and annotation of representative corpora of real-world stimuli, faces logistical challenges, but these can be overcome by solutions that are deployed in the later stages of integrative experiment design.

The integrative approach advocated by Almaatouq et al. starts with mapping a research field onto an *n*-dimensional design space that defines the universe of relevant experiments – what they call "research cartography" (target article, sect. 3.1 para. 2). They suggest that the design space's dimensions can be extracted from available taxonomies, prior experimental research, and practical experience. However, as they acknowledge, this approach is vulnerable to unknown unknowns: Taxonomies, prior experiments, and practical experience may all fail to identify important dimensions which should be included in the design space.

Here, we focus on one way of identifying unknown unknowns: High-throughput natural description. This approach may help research cartographers to uncover missing dimensions of the research design space, at a cost comparable to the later stages of the integrative experiment design.

To appreciate the value of high-throughput natural description, consider cases where researchers noticed a discrepancy between the experimental stimuli and the naturalistic variation of these stimuli. For instance, Schutz and Gillard (2020) showed that many experiments studying nonspeech auditory perception used flat tones as stimuli, despite the fact that such tones are unrealistic: Their content lacks dynamic changes found in the temporal structure of naturalistic sounds. Experiments that included such naturalistic content made novel discoveries about the auditory system. For example, a study of audiovisual integration showed that tones with a temporal structure similar to impact sounds, like the sound of a xylophone, but not flat tones, which lack temporal variation, were reliably integrated with visual information when participants judged tone duration (Schutz & Kubovy, 2009).

Similarly, Dawel, Miller, Horsburgh, and Ford (2021) and Barrett, Adolphs, Marsella, Martinez, and Pollak (2019) showed that many experiments studying face perception used highly standardised and posed facial configurations which are not representative of the real-world variation in facial configurations. When naturalistic facial configurations are used in experiments, reported findings differ from previous results. For example, using naturalistic facial stimuli, Sutherland et al. (2013) found that facial first impressions have three underlying dimensions (trustworthiness, dominance, and youthfulness/attractiveness) instead of just two (trustworthiness and dominance), as previously reported when standardised facial stimuli were used (Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008).

In these examples, researchers noticed and resolved some discrepancy between the variation of experimental and real-world stimuli. Such an approach, while useful, does not completely solve the problem of unknown unknowns. This is because there may be many more real-world variations in stimuli that could update one's understanding of a phenomenon, if they were introduced in experimental designs. However, a researcher cannot identify them unless they have a thorough description of real-world variation.

One solution to this issue is "high-throughput natural description": *The systematic collection and annotation of large, representative corpora of real-world stimuli to identify unknown unknowns.*

An example in the field of emotion perception demonstrates the value of this approach. By collecting and annotating 7 million pictures of faces and 10,000 hours of filmed video from the internet, Srinivasan and Martinez (2018) discovered that the emotion-category labels of disgust, anger, sadness, and happiness are associated with 1, 5, 5, and 17 "distinct" facial configurations, respectively. Such variation in the range of facial configurations conveying different emotions was an unknown unknown in the research cartography of emotion perception, and studies investigating responses to facial configurations expressing certain emotion categories have yet to investigate responses to the entirety of the observed variation, to the best of our knowledge (Barrett et al., 2019). Thus, high-throughput natural description can aid in defining the design space of relevant experiments via the identification of unknown unknowns.

However, this solution is not an easy fix to the problem of unknown unknowns. Large-scale naturalistic observation is logistically challenging. Obtaining 7 million images of faces from the internet is in itself difficult, but the difficulty ramps up if researchers wish to obtain a sample of faces from more diverse sources. Furthermore, large-scale annotation can be as challenging as large-scale naturalistic observation. For example, creating a corpus of 7 million faces that is useful for answering different research questions requires annotating the images for meaningful dimensions. Coding action units (specific facial muscle movements) manually via human annotators in these images can require expertise, or can take years when the dataset is extremely large (Benitez-Quiroz, Srinivasan, & Martinez, 2016; Srinivasan & Martinez, 2018). Furthermore, the pool of annotators must itself be (very) large, not only to deal with the size of the corpus, but also to identify relevant individual and cultural variations in the way coders perceive the dimensionality of the stimuli.

In sum, while high-throughput natural description aids in the identification of unknown unknowns of a research design space, it introduces significant logistical challenges. However, these challenges can be surmounted via a combination of *mass collaboration*, *automation* (a use case is already present in the aforementioned emotion perception example where Srinivasan & Martinez, 2018, use a computer vision algorithm to annotate action units in the internet images; Benitez-Quiroz et al., 2016; Yitzhak et al., 2017), *citizen science* (Awad et al., 2018, 2020; Hilton & Mehr, 2021), and *gamification* (Long, Simson, Buxó-Lugo, Watson, & Mehr, 2023). In fact, Almaatouq et al. already propose that these aforementioned solutions could be deployed in the later stages of the integrative experiment design

Nonetheless, the application of these solutions for executing high-throughput natural description should not be ignored, as they amplify concerns about the up-front costs and inclusivity of the integrative approach. Few research groups may have the resources to implement an integrative experiment design, and fewer groups still may be able to solve its unknown unknowns problem during the research cartography stage. While we are enthusiastic about the ideas in the target article, we believe it is necessary to be explicit and constructive about the requirements of an integrative experiment design approach.

Acknowledgments. T. K. would like to thank Dr. Julie Grèzes for briefly discussing the current state of the face perception and social cognition literature.

Financial support. S. A. M. is supported by NIH DP5OD024566. J.-F. B. acknowledges support from grants ANR-19-PI3A-0004, ANR-17-EURE-0010, and the research foundation TSE-Partnership.

Competing interest. None.

References

Awad, E., Dsouza, S., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing moral machines. *Communications of the ACM*, *63*(3), 48–55.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., … Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*, 59–64.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2016). EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In 2016 IEEE Conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA (pp. 5562–5570). https://doi.org/10.1109/CVPR.2016.600

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2021). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*, *54*(4), 1889–1901. https://doi.org/10.3758/s13428-021-01705-3

Hilton, C., & Mehr, S. (2021). Citizen science can help to alleviate the generalizability crisis.*45*, e21.

Long, B., Simson, J., Buxó-Lugo, A., Watson, D. G., & Mehr, S. A. (2023). How games can make behavioural science better. *Nature*, *613*(7944), 433–436.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Schutz, M., & Gillard, J. (2020). On the generalization of tones: A detailed exploration of non-speech auditory perception stimuli. *Scientific Reports*, *10*(1), 9520. https://doi.org/10.1038/s41598-020-63132-2

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1791.

Srinivasan, R., & Martinez, A. M. (2018). Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild. *IEEE Transactions on Affective Computing*, *12*(3), 707–721.

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Yitzhak, N., Giladi, N., Gurevich, T., Messinger, D. S., Prince, E. B., Martin, K., & Aviezer, H. (2017). Gently does it: Humans outperform a software classifier in recognizing subtle, nonstereotypical facial expressions. *Emotion*, *17*(8), 1187–1198. https://doi.org/10.1037/emo0000287

Against naïve induction from experimental data

David Kellen[a]*, Gregory E. Cox[b], Chris Donkin[c], John C. Dunn[d] and Richard M. Shiffrin[e]

[a]Department of Psychology, Syracuse University, Syracuse, NY, USA; [b]Department of Psychology, College of Arts and Sciences, University at Albany, State University of New York, Albany, NY, USA; [c]Department of Psychology, Ludwig Maximilian University of Munich, München, Germany; [d]Department of Psychology, University of Western Australia, Perth, WA, Australia and [e]Psychological and Brain Sciences Department, Indiana University, Bloomington, IN, USA
davekellen@gmail.com
gecox@albany.edu
christopher.donkin@gmail.com
john.dunn@uwa.edu.au
shiffrin@indiana.edu

*Corresponding author.

doi:10.1017/S0140525X2300211X, e51

Abstract

This commentary argues against the indictment of current experimental practices such as piecemeal testing, and the proposed integrated experiment design (IED) approach, which we see as yet another attempt at automating scientific thinking. We identify a number of undesirable features of IED that lead us to believe that its broad application will hinder scientific progress.

After so many years observing the prosecution of *p*-values and everyday laboratory life, we are pleased to see a growing number of researchers turning their attention to critical matters such as theory development and experimentation (e.g., Proulx & Morey, 2021). But as we transition into these important new debates, it is crucial to avoid past intellectual excesses. In particular, we note a tendency to embrace passive technological solutions to problems of scientific inference and discovery that make little room for the kind of active theory building and critical thinking that in fact result in meaningful scientific advances (see Singmann et al., 2023). In this vein, we wish to express serious reservations regarding Almaatouq et al.'s critique.

The observation of puzzling, incongruent, and incommensurate results across studies is a common affair in the experimental sciences (see Chang, 2004; Galison, 1987; Hacking, 1983). Indeed, one of the central roles of experimentation is to "create, produce, refine and stabilize phenomena" (Hacking, 1983, p. 229), which is achieved through an iterative process that includes the ongoing improvement of experimental apparati (see Chang, 2004; Trendler, 2009) and relevant variables (Jantzen, 2021). This process was discussed long ago by Maxwell (1890/1965), who described it as removing the influence of "disturbing agents" from a "field of investigation."