Received: 31 October 2022 | Accepted: 1 December 2022

DOI: 10.1111/bjop.12622

the british psychological society
promoting excellence in psychology

COMMENTARY

Generative models for visualizing idiosyncratic impressions

Alexander Todorov | Stefan Uddenberg | Daniel Albohn

The University of Chicago Booth School of Business, Chicago, Illinois, USA

Correspondence
Alexander Todorov, The University of Chicago Booth School of Business, Chicago, IL, USA.
Email: alexander.todorov@chicagobooth.edu

Abstract

In their comprehensive review of research on impressions from faces, Sutherland and Young (this issue) highlight both the remarkable progress and the many challenges facing the field. We focus on two of the challenges: the need for generative, powerful models of impressions and the idiosyncratic nature of complex impressions.

KEYWORDS
data-driven methods, faces, idiosyncratic differences, impressions

BUILDING OF POWERFUL GENERATIVE MODELS

As outlined in the target article, data-driven methods have led to tremendous progress in identifying cues for various psychological impressions. Importantly, data-driven methods are not constrained by researchers' prior heuristics or biases. However, they are susceptible to certain constraints. The most important of these is the nature of the stimuli or the stimulus variation of the input space. These data-driven methods are akin to reverse correlation methods, in which random variations in the stimuli (e.g., facial images) are classified as a function of behaviour (e.g. judgement of facial images). Broadly, these data-driven methods fall into two classes: psychophysical reverse correlation methods—in which participants judge facial stimuli altered by visual noise—and face space reverse correlation methods—in which participants judge faces randomly generated by a statistical multi-dimensional face space (Todorov et al., 2011). A third technique dates back to Galton's composite photography and consists of morphing facial images selected to exemplify particular categories of stimuli (e.g. an intelligent-looking face; Sutherland et al., 2013). The objective of these techniques is to identify some systematic variation in the stimuli that predicts judgements. This systematic variation is then interpreted as comprising the set of cues driving the specific judgement.

Dozens of impression models have been created and validated using these techniques (Brinkman et al., 2017; Sutherland et al., 2013; Todorov & Oh, 2021; Vernon et al., 2014; Walker & Vetter, 2016). But each technique suffers from its own limitations. The images in psychophysical reverse correlation studies are noisy by design and, as a result, the final classification images (e.g. morphs of blurry images selected to look more intelligent) are also noisy, rarely revealing subtle differences in cues. Worse, these models often fail at the level of individual participants, a failure related to the second challenge facing the field named earlier. Face space-based reverse correlation studies use images of synthetic faces that lack the realism and diversity of real faces. Moreover, the original face space models were constructed from laser scans of only a few hundred (mostly white) faces, which are not representative of the diversity of human faces. Similar concerns apply to morphing techniques, which are limited by the set of original faces and can produce artefacts in the exemplar images such as warping or the appearance of 'smoothed' skin and hair.

The rapid development of deep machine learning methods, specifically generative adversarial networks, offers a potential solution to the challenge of limited stimulus variation. By using massive data sets of images, these methods are capable of modelling the representation of faces (Karras et al., 2018, 2020). Although the resulting representations are less readily interpretable (unlike the older models; O'Toole et al., 2018), in principle these models are capable of generating an unlimited number of hyper-realistic faces that can capture the diversity of human faces. Capitalizing on one of these models (StyleGAN; Karras et al., 2020), with our collaborators we created more than 30 models of impressions, ranging from judgements of relatively unambiguous attributes such as perceived masculinity to judgements of completely subjective attributes such as perceived familiarity (Peterson et al., 2022). The resulting models can also be applied to existing real faces, once these faces are encoded in the latent face space. These new generative face models overcome many of the limitations of prior models. But even the new models are not entirely free of stimulus-based constraints. Although StyleGAN was trained on a massive set of 70,000 face photographs, it is not obvious that the resulting space is fully representative of the diversity of faces, especially with respect to the range of possible emotional expressions humans can produce.

The second constraint of data-driven methods is that the resulting models would always reflect the impressions of the specific group of raters. After all, these methods are designed to reveal the facial representations of these raters. To the extent that the majority of raters happened to be white, for example, the resulting models would simply reflect the stereotypes of this particular group. Thus, even though it is straightforward to create models of the appearance of faces representing different ethnicities (Peterson et al., 2022), these models would not reflect the actual diversity of the groups but rather the raters' visual stereotypes of these groups.

Finally, building good models requires much larger samples of both faces and raters than in typical psychological studies. In our studies, we used ~1000 faces (where each face was rated by at least 30 distinct raters) to build the models, but simulations showed that our models could be further improved by including more faces. This was particularly the case for judgements of complex perceived attributes (e.g. 'electable', 'trustworthy') in contrast to simple attributes (e.g. 'hair colour'). This was also true for the number of raters per face—more raters lead to more variance explained in the models.

Data-driven methods almost always find reliable facial cues predictive of a given set of judgements. But these cues are always a function of the specific face stimuli and raters involved. To find out the extent to which these cues generalize across faces and raters, we would need multiple generative models built from large data sets in terms of both the number of faces and raters. The potential convergence of such models would be an indication of the universality of facial cues that drive complex (and subjective) impressions.

THE DEEPLY IDIOSYNCRATIC NATURE OF COMPLEX IMPRESSIONS

Almost all models of impressions, including the ones discussed above, are models of judgements aggregated across raters. This procedure masks massive individual differences in complex impressions. Sutherland and Young (2022) discuss this challenge, but we believe that this is one of the most important, most underappreciated, and least developed areas of research on first impressions.

The first researcher to systematically study the idiosyncratic nature of impressions was Hönekopp (2006). Using variance decomposition analysis as applied to judgements of attractiveness, he showed that idiosyncratic differences account for about 50% of the meaningful variance of these judgements. Upon reading his work, one of the authors (AT) assumed that Hönecopp might have overestimated the share of idiosyncratic variance. This assumption was based on the observation that inter-rater agreement is higher when it is computed from the aggregated judgements of the raters (e.g. an average of the rater's multiple judgements of the same faces in contrast to a single judgement). Because the raters in Hönecopp's studies made only two judgements (of the same stimuli), AT reasoned that much of the error variance is added to the idiosyncratic variance. AT was wrong. Although multiple repeated measurements increase inter-rater agreement, the increase in the intra-rater agreement is much steeper. As a result, multiple (and more reliable) measurements, if anything, increase the estimates of the relative share of the idiosyncratic variance (Martinez et al., 2020). This paper (Martinez et al., 2020), in addition to confirming Hönecopp's insights, extended the methods to multiple visual stimuli and multiple judgements. The paper also provides methodological guidelines about the number of raters, stimuli, and repeated measurements and how these affect the precision of the variance estimates.

Hehman et al. (2017) were the first to apply the variance partitioning reasoning to a broad range of impressions. Unfortunately, most of the studies in this paper did not include repeated measurements. The latter is essential to estimate the most interesting component of idiosyncratic differences (i.e. the interaction of perceiver and stimulus; the other component is the perceiver's variance) and to remove the error variance from the meaningful variance. Nevertheless, their estimates are comparable to other more recent studies (Albohn et al., 2022; Hester et al., 2021; Martinez et al., 2020). In cases of complex judgements such as perceived trustworthiness and competence, the idiosyncratic variance trumps shared variance. Yet the latter is what is modelled in studies on first impressions, as explained in the previous section.

All of the studies on estimating shared and idiosyncratic variance so far have been descriptive: documenting the existence of an important phenomenon. But we know very little about the variables predicting the relative shares of these variances and the mechanisms leading to stable idiosyncratic differences. To be fair, such mechanisms were discussed in the target article. However, to the best of our knowledge, we do not know of any study in which the variance of impressions is partitioned and then its components are predicted by a set of variables postulated by the researchers. In fact, researchers rarely include repeated measures in their studies. Yet these measures are critical for advancing the study of idiosyncratic differences in impressions.

If documenting the importance of idiosyncratic variance is the first step in the study of idiosyncratic differences in impressions, one of the final steps is building models of idiosyncratic representations of impressions. Recently, capitalizing on the power of generative face models and borrowing procedures from psychophysical reverse correlation, we have proposed and illustrated methods for building such representations (Albohn et al., 2022). But many questions remain unresolved. For one, it is not obvious how to analytically relate statistical studies on variance decomposition and modelling studies of impressions. Specifying this relation is critical to formally test different mechanisms within the same computational framework and to study the predictive utility of idiosyncratic models.

AUTHOR CONTRIBUTIONS

Alexander Todorov: Conceptualization; writing – original draft. **Stefan Uddenberg:** Conceptualization; writing – review and editing. **Daniel Albohn:** Conceptualization; writing – review and editing.

CONFLICT OF INTEREST

All authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

There are no data associated with this paper.

ORCID

Alexander Todorov https://orcid.org/0000-0003-1271-6113

REFERENCES

Albohn, D. N., Uddenberg, S., & Todorov, A. (2022). A data-driven, hyper-realistic method for visualizing individual mental representations of faces. *Frontiers in Psychology*, *13*, 997498.

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*, 333–361.

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*, 513–529.

Hester, N., Xie, S. Y., & Hehman, E. (2021). Little between-region and between-country variance when forming impressions of others. *Psychological Science*, *32*, 1907–1917.

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 199–209.

Karras, T., Laine, S., & Aila, T. (IEEE, 2018). A style-based generator architecture for generative adversarial networks. In *CVF conference on computer vision and pattern recognition (CVPR)* (pp. 4396–4405). Institute of Electric and Electronics Engineers.

Karras, T., et al. (IEEE, 2020). Analyzing and improving the image quality of StyleGAN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8110–8119). Institute of Electric and Electronics Engineers.

Martinez, J. E., Funk, F., & Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to judgment. *Behavior Research Methods*, *52*, 1428–1444.

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in Cognitive Sciences*, *22*, 794–809.

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences of the USA*, *119*(17), e2115228119.

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

Sutherland, C. A. M., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*, *113*, 1056–1078.

Todorov, A., Dotsch, R., Wigboldus, D., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*, *5*, 775–791.

Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. *Advances in Experimental Social Psychology*, *63*, 189–246.

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS*, *111*(32), E3353–E3361.

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived big two and big five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, *110*(4), 609–624.

How to cite this article: Todorov, A., Uddenberg, S., & Albohn, D. (2023). Generative models for visualizing idiosyncratic impressions. *British Journal of Psychology*, *114*, 511–514. https://doi.org/10.1111/bjop.12622