References

1. Smaers, J.B. *et al.* (2017) Exceptional evolutionary expansion of prefrontal cortex in great apes and humans. *Curr. Biol.* 27, 714–720
2. Croxson, P.L. *et al.* (2017) Structural variability across the primate brain: a cross-species comparison. *Cereb. Cortex* Published online October 13, 2017. http://dx.doi.org/10.1093/cercor/bhx244
3. Phillips, K.A. and Sherwood, C.C. (2008) Cortical development in brown capuchin monkeys: a structural MRI study. *Neuroimage* 43, 657–664
4. Fears, S.C. *et al.* (2011) Anatomic brain asymmetry in vervet monkeys. *PLoS One* 6, e28243
5. Love, S.A. *et al.* (2016) The average baboon brain: MRI templates and tissue probability maps from 89 individuals. *Neuroimage* 132, 526–533
6. Atkinson, E.G. *et al.* (2015) Cortical folding of the primate brain: an interdisciplinary examination of the genetic architecture, modularity, and evolvability of a significant neurological trait in pedigreed baboons (genus *Papio*). *Genetics* 200, 651–665
7. Gómez-Robles, A. *et al.* (2015) Relaxed genetic control of cortical organization in human brains compared with chimpanzees. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14799–14804
8. Han, S. and Ma, Y. (2015) A culture–behavior–brain loop model of human development. *Trends Cogn. Sci.* 19, 666–676
9. Sherwood, C.C. and Gómez-Robles, A. (2017) Brain plasticity and human evolution. *Annu. Rev. Anthropol.* 46, 399–419
10. Van Essen, D.C. and Dierker, D.L. (2007) Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* 56, 209–225

Forum

A Dynamic Structure of Social Trait Space

Ryan M. Stolier,[1,*] Eric Hehman,[2] and Jonathan B. Freeman[1,*]

Facial appearance evokes robust impressions of other people's personality traits. Recent research suggests that the trait space arising from face-based impressions shifts due to context and social cognitive factors. We suggest a novel framework in which multiple bottom-up and top-down processes mutually determine a dynamic rather than fixed trait space.

Humans form a variety of impressions of others based on facial appearance. Such face-based trait judgments (e.g., friendly) are argued to be primarily structured by two relatively independent dimensions, trustworthiness and dominance, such that all possible trait judgments tend to fall along these two primary axes [1]. These fundamental dimensions are often assumed to reflect a fixed and universal architecture, sharing a kinship with other models of social perception in which the assumption has been made explicit (e.g., [2]).

Much research has focused on the role of bottom-up stimulus attributes, richly specifying sets of highly consistent facial features that drive particular judgments. Current explanations of the dimensional space of judgments (i.e., 'trait space') focus on inherent similarity in bottom-up features driving similarity in inferred traits. For instance, subtle emotional cues (e.g., mouth shape resembling a smile) convey both trustworthiness and extroversion, whereas physical strength cues (e.g., a heavy brow) convey both dominance and aggression [1]. Such explanations therefore posit a relatively fixed account of how trait judgments interrelate from shared features. While such approaches have been highly valuable, recent research has documented several cases in which the structure of trait space shifts (i.e., the correlations among trait judgments change). This leads the fundamental dimensions to change in their relationship or may even result in the emergence of new dimensions altogether. Although prior research has demonstrated top-down effects in individual trait judgments [3], top-down contributions are not central to current models of trait evaluation and trait space. Here we describe a new perspective, using quantitative techniques borrowed from systems neuroscience, to illuminate how multiple bottom-up facial features and top-down social cognitive processes together shape a dynamic trait space.

Trait space has recently been shown to shift in a variety of contexts in which stereotypes, motives, or group processes are likely to play a role. Stereotypes have been found to exert a notable effect where target stereotypes shift the evaluation of specific traits. For instance, females are more positively evaluated when their personalities are submissive rather than dominant, due to stereotypical expectancies that women need protection and coddling (i.e., 'benevolent sexism'). Accordingly, trustworthiness judgments are more negatively related to perceived dominance in female compared with male faces [4]. Similarly, when faces of older adults are evaluated, facial dominance is considerably less tethered to trustworthiness [5], as stereotypes of physical frailty buffer against negative implications of appearing dominant and hostile. Cultural factors may also shift trait space. For example, perceived trustworthiness depends more on facial typicality cues for own-culture faces, due to awareness of own-culture norms in facial appearance, while depending more on attractiveness cues for other-culture faces [6].

Social motivations, like the desire to evaluate close others more positively than distant others, also shift trait space. For example, people may construe dominance traits as positive for close and trusted others but as threatening and negative for distant others (i.e., evaluation of strength in a friend vs a foe). Accordingly, dominance and trustworthiness are positively correlated when judging close and admired others (e.g., [7]) but negatively correlated when judging unfamiliar and outgroup others (e.g., [2]). Recent work further suggests that close as opposed to distant others may be represented in higher-dimensional spaces more generally, perhaps due to more complex representations of familiar personalities [8]. Together these recent findings suggest that top-down social

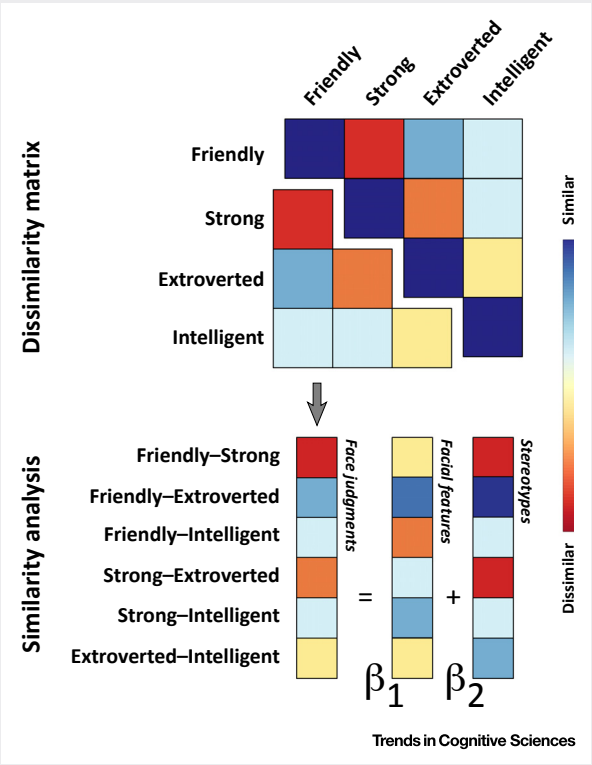

Figure I. Representational Similarity Analysis

Box 1. Representational similarity analysis

RSA provides an intuitive analytic framework to quantify how one multidimensional space can be constructed by the weighted integration of others. Initially developed in systems neuroscience to branch levels of analysis (brain, behavior, and computational models), RSA provides a powerful tool to integrate distinct yet related levels of measurement (e.g., perceptual stimuli similarities as predicted by brain data, behavioral tasks, and computational models). For our purposes here, RSA measures how the similarity structure of variables (e.g., traits) measured in one context (e.g., judgments) may be predicted by their similarity structure in other contexts (e.g., facial features, stereotypes). This analysis is simple and intuitive in that it predicts one vector of variable pair-wise similarities with another (unique values under the diagonals of symmetric similarity matrices), allowing the use of familiar analysis techniques that measure variance explained in one variable by others (e.g., bivariate correlation, multiple regression). Here an example of multiple regression RSA is depicted (Figure I) in which unique pair-wise similarities in face judgments are predicted by a linear combination of analogous pair-wise similarities in facial features as well as stereotypes. We should highlight the flexibility of this method to test various kinds of hypotheses not illustrated here. Extensive detail on this method can be found elsewhere [10].

cognitive factors may shift the dimensional space underlying face-based trait perception and thus how individuals' faces are evaluated within it.

Here we suggest a theoretical framework in which trait space reflects the integration of multiple bottom-up and top-down representational spaces. In this framework, we consider trait space as a network of trait concepts, with all trait-concept pairs associated via weighted connections (e.g., association of warmth and power) (see also [9]). One means to characterize such a trait space utilizes the representational similarity analysis (RSA) technique from systems neuroscience (Box 1) [10]. As such, trait space may be characterized in a similarity-matrix form in which each cell is the pair-wise weighted connection or 'similarity' of two trait concepts. Consider a set of eight trait judgments with an 8×8 matrix capturing the similarity (association) of any given pair of traits with respect to judgments (Figure 1). Each representational space describes the similarity of trait pairs on the basis of a specific factor (e.g., similarity of warmth and power due to facial features or due to stereotypes). Our theoretical premise is that the trait space of judgments reflects an integration of not only bottom-up but also top-down representational spaces. Thus, quantitatively, trait space should reflect a linear combination of such spaces. In the example of Figure 1, each trait-pair similarity in judgments should reflect a combination of the corresponding similarity in both facial features and stereotypes.

For example, in the context of female targets, consider the relationship between 'warm' and 'powerful' in facial features and stereotypes (Figure 1). First, the relation of warmth and power judgments as driven by facial features could be modeled, reflected in the facial-feature similarity matrix. For example, this could capture the overlap in which facial metrics (e.g., brow width) correlate with both warmth and power judgments. Second, the stereotypically negative association between warmth and power for women [4] can also be measured, reflected in a negative relationship between warmth and power in the stereotype similarity matrix. For example, this could capture the extent to which people believe warmth and power personality traits correlate for women. In this framework the prediction is that the correlation of warmth and power in final perceptual judgments reflects an integration of the two traits' relationship in both the facial-features and the stereotype matrices (e.g., see the linear summation of warm–

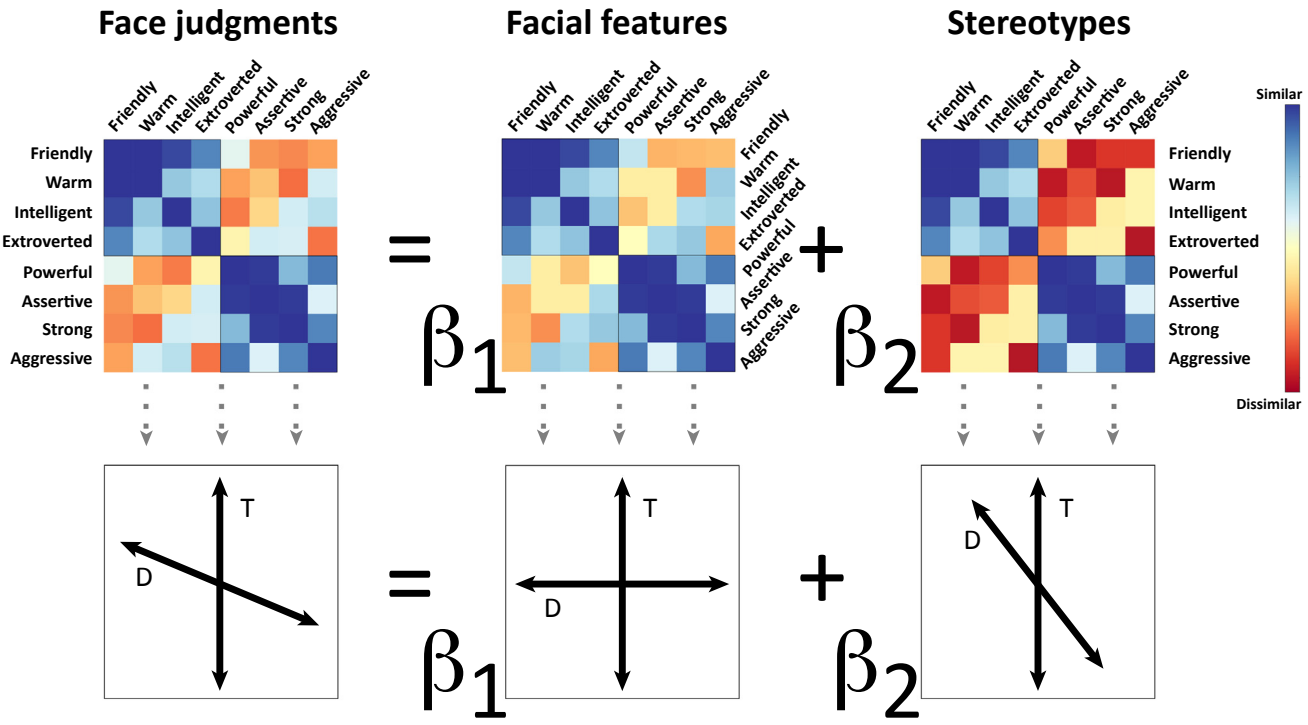

Figure 1. Schematic Illustration of Our Proposed Framework. An example trait space of face judgments (i.e., trait-pair similarities) for female targets is depicted as a linear function of multiple contributing factors – in this case, facial features and stereotypes – using hypothetical data. Each matrix holds the trait pair-wise similarities (associations) on the basis of different data: their perceptual judgment, intrinsic facial features, and stereotypical associations (e.g., gender stereotypes). For each, trait pairs are measured and found to be more similar (blue) or dissimilar (red) (e.g., using correlation). In this example two clusters emerge in the matrices (blue-boxed values), translating to unique dimensions ('trustworthiness' and 'dominance', as could be derived by dimensionality-reduction techniques) that are loosely anticorrelated in face judgments, as argued by current trait space models [1]. Below each similarity matrix are the hypothetical corresponding derived dimensions (e.g., principal components): trustworthiness (T) and dominance (D). In this example our framework argues that face-judgment pair-wise similarities can be predicted by the linear combination of their corresponding similarities in inherent facial features as well as stereotypes. In turn, this means that the two principal dimensions (above, T and D) may shift due to integration of other representational spaces. Thus, here, their slightly negative relationship in judgments for female targets, as recently observed [4], arises due to the linear combination of a weak relationship in facial features with a strong negative relationship from gender stereotypes.

powerful similarity cells in Figure 1). Note that this framework does not operate on a single example pair, but rather, comprehensively assesses how the entire system of relations among all trait pairs (trait space) reflects an integration across multiple systems of trait relations. In turn, this can also account for how principal dimensions structuring the space (e.g., trustworthiness and dominance) may shift due to the impact of other representational spaces (Figure 1).

More generally, any number of representational spaces may be measured and implemented, permitting an understanding of how perceptual features may be integrated with countless social cognitive factors (e.g., stereotypes, context, familiarity, person knowledge, emotion, motivation). Indeed, RSA was recently applied to tease apart the contributions of facial features and stereotypical associations in neural representations of gender, race, and emotion categories [11]. As such, this framework provides a flexible and powerful quantitative means to explore our theoretical premise that a comprehensive system of relations, shaped by both bottom-up and top-down factors, drives an individual's trait perceptions. This framework also provides an opportunity to explore perceptual spaces in full, without necessarily requiring data-reduction techniques that simplify the space to some putative set of core dimensions. It provides a means to formalize predictions of how specific factors determine trait space, but of course it complements numerous other quantitative methods that may usefully apply here.

Our perspective may be valuable for many other popular two-dimensional models in social cognition and cognitive science more generally (e.g., [1,2]). One question of growing interest is when such classic two-dimensional models fail to adequately

account for the data relative to higher-dimensional models. For instance, lower-dimensional spaces may be a consequence of the experimental task or the specific stimulus set, obscuring higher-dimensional spaces. As one example, attractiveness is related to trustworthiness in a sample of faces with a limited age range [1], but related to a unique dimension of youthfulness when faces vary by age [12]. Such questions are uniquely well suited to our proposed framework, as modeling the multiple representational spaces driving judgments opens the door to the assessment and prediction of which factors may constrict or expand trait space. Finally, face-based trait perceptions are consequential, impacting outcomes like criminal sentences or elections [3]. If different contexts shift the relationship between trait judgments, we may more easily identify groups vulnerable to certain impression-formation patterns and crucial moments in which limited personality information could bias wider personality judgments and stereotypes.

In short, a growing body of studies is making clear that, rather than any fixed architecture of trait judgments, there are specific trait spaces that arise depending on different contexts and targets or particular social cognitive processes. We hope that this framework may prove valuable as researchers aim to better understand the dynamic nature of social trait space.

Acknowledgments

This work was funded in part by National Science Foundation research grant NSF-BCS-1654731 (J.B.F).

[1]New York University, New York, NY, USA
[2]Ryerson University, Toronto, Canada

*Correspondence:
rystoli@nyu.edu (R.M. Stolier) and
jon.freeman@nyu.edu (J.B. Freeman).

http://dx.doi.org/10.1016/j.tics.2017.12.003

References

1. Oosterhof, N.N. and Todorov, A. (2008) The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11087–11092
2. Cuddy, A.J.C. *et al.* (2009) Stereotype content model across cultures: towards universal similarities and some differences. *Br. J. Soc. Psychol.* 48, 1–33
3. Todorov, A. *et al.* (2015) Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annu. Rev. Psychol.* 66, 519–545
4. Sutherland, C.A. *et al.* (2015) Face gender and stereotypicality influence facial trait evaluation: counter-stereotypical female faces are negatively evaluated. *Br. J. Psychol.* 106, 186–208
5. Hehman, E. *et al.* (2014) The face–time continuum: lifespan changes in facial width-to-height ratio impact aging-associated perceptions. *Pers. Soc. Psychol. Bull.* 40, 1624–1636
6. Sofer, C. *et al.* (2017) For your local eyes only: culture-specific face typicality influences perceptions of trustworthiness. *Perception* 46, 914–928
7. Kraft-Todd, G.T. *et al.* (2017) Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PLoS One* 12, e0177758
8. Thornton, M.A. and Mitchell, J.P. (2017) Theories of person perception predict patterns of neural activity during mentalizing. *Cereb. Cortex* Published online August 22, 2017. http://dx.doi.org/10.1093/cercor/bhx216
9. Freeman, J.B. and Ambady, N. (2011) A dynamic interactive theory of person construal. *Psychol. Rev.* 118, 247–279
10. Kriegeskorte, N. *et al.* (2008) Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4
11. Stolier, R.M. and Freeman, J.B. (2016) Neural pattern similarity reveals the inherent intersection of social categories. *Nat. Neurosci.* 19, 795–797
12. Vernon, R.J. *et al.* (2014) Modeling first impressions from highly variable facial images. *Proc. Natl. Acad. Sci. U. S. A.* 111, E3353–E3361