Generative Neural Networks for Experimental Manipulation of Complex Psychological Impressions in Face Perception Research and Beyond

Adam Sobieszek
Department of Psychology, University of Warsaw, Warsaw, Poland

Online

Introduction

The work aims to generalize the computational models of psychological impressions, such as trustworthiness, pioneered in face perception, to the rich domain representations learned by neural networks. Such models may then be used to manipulate these impressions.

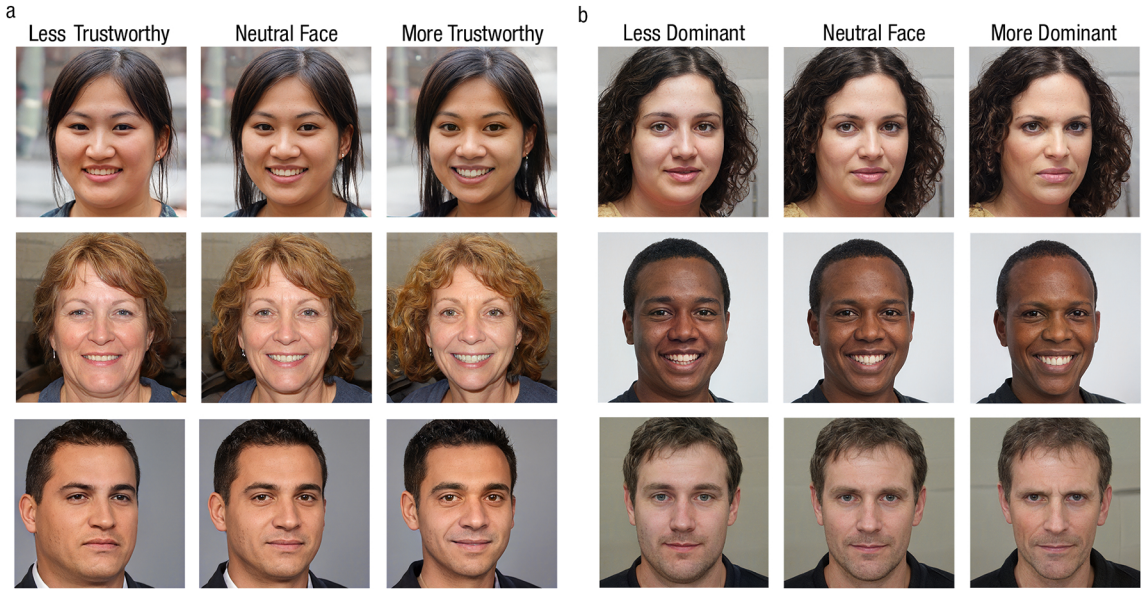
The method produces real-looking stimuli for psychological studies that can be minimally manipulated to elicit higher or lower levels of a particular impression.

Results of an application of this method to the manipulation of trustworthiness and dominance of faces are presented.

Method

GANs are neural networks that learn to generate new stimuli similar to a dataset of training examples. They also learn a vector representation of the data, called the latent space, where each point corresponds to a stimuli they can generate.

We may construct a model of an impression on this latent space by asking participants to rate stimuli randomly drawn from it. We fit a function that predicts ratings based on position in the latent space. The gradient of this function is the direction in which we should manipulate the stimuli.



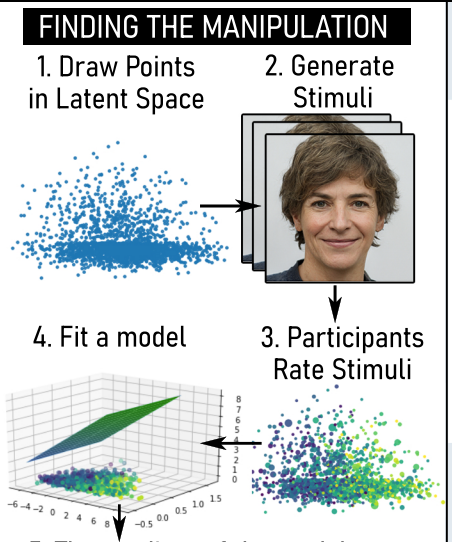


Figure 1. Example manipulations obtained with the present method.

Validation

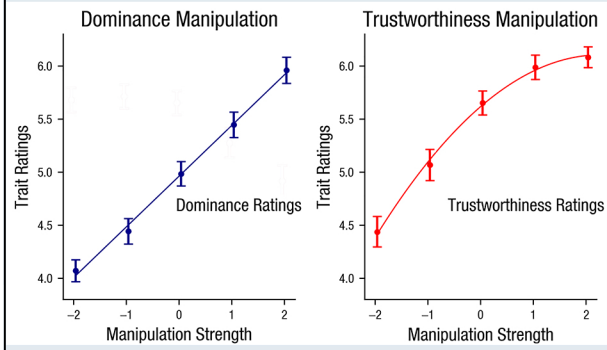


Figure 2. Results of a validation study, a between-subject design with 5 levels of manipulation was used. Participants rated 45 faces on a 9-point scale. Mean ratings in each experimental condition are presented.

Conclusion

The method allows for generation of arbitrarily many stimuli manipulated on a modelled impression. The manipulation is valid, as it is data-driven and based on human judgements. The manipulation also has a greater ability to institute controlled variables as it does not require picking them explicitly. Instead, it exploits the fact that the network learnt the factors of variation in the data, such that when we move down the gradient of the impressions all other features are controlled, because they are associated with directions that are orthogonal in latent space.