Cognition 158 (2017) 81–89

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Physical attraction to reliable, low variability nervous systems: Reaction time variability predicts attractiveness

Emily E. Butler [a,b], Christopher W.N. Saville [b,c], Robert Ward [b], Richard Ramsey [b,*]

[a] Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[b] Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, Gwynedd, Wales, UK
[c] University Hospital for Child and Adolescent Psychiatry, University of Freiburg, Germany

ARTICLE INFO

Article history:
Received 3 December 2015
Revised 11 October 2016
Accepted 25 October 2016

Keywords:
Face perception
Attractiveness
Reaction time variability
Central nervous system

ABSTRACT

The human face cues a range of important fitness information, which guides mate selection towards desirable others. Given humans' high investment in the central nervous system (CNS), cues to CNS function should be especially important in social selection. We tested if facial attractiveness preferences are sensitive to the reliability of human nervous system function. Several decades of research suggest an operational measure for CNS reliability is reaction time variability, which is measured by standard deviation of reaction times across trials. Across two experiments, we show that low reaction time variability is associated with facial attractiveness. Moreover, variability in performance made a unique contribution to attractiveness judgements above and beyond both physical health and sex-typicality judgements, which have previously been associated with perceptions of attractiveness. In a third experiment, we empirically estimated the distribution of attractiveness preferences expected by chance and show that the size and direction of our results in Experiments 1 and 2 are statistically unlikely without reference to reaction time variability. We conclude that an operating characteristic of the human nervous system, reliability of information processing, is signalled to others through facial appearance.

1. Introduction

Theories of mate selection emphasise the role of attractiveness preferences for guiding mate-choice towards high fitness partners. Specifically, traits that are associated with high fitness should be attractive to potential mates because they offer advantages to a partner as well as to future offspring (Gangestad & Scheyd, 2005; Rhodes, 2006). In humans, facial attractiveness preferences have been repeatedly shown to reflect aspects of mate quality. For example, facial symmetry is attractive and indicative of developmental stability and resilience (Simmons, Rhodes, Peters, & Koehler, 2004) and certain levels of facial colouration are attractive and denote healthy blood oxygenation (Stephen, Coetzee, Smith, & Perrett, 2009). The human face thus reflects a range of important fitness information. Given humans' high investment in the central nervous system (CNS), we would predict that cues to CNS function would be especially important in mate selection. To date, however, there is no evidence that facial appearance specifically reflects the reliability of the CNS; that is, the degree to which the nervous system functions in a consistent manner. Although consistency of behaviour would cue CNS reliability, appearance cues would offer the advantages of rapid assessment for observers as well as rapid signalling for high-fitness senders. We therefore tested if facial attractiveness preferences are sensitive to the reliability of human nervous system function.

A reliable information processor will produce relatively invariant outputs for a specified input (Shannon, 1948). One important constraint on the reliability of an information processor is the amount of endogenous noise. Endogenous noise can be defined as the amount of unpredictable fluctuation across processing operations within a system. Increasing amounts of endogenous noise eventually become an enemy of reliable information processing (Faisal, Selen, & Wolpert, 2008; Shannon, 1948). In other words, given a repeated input, low noise systems will be reliable, in the sense of producing invariant output. In contrast, high noise systems will produce more variable outputs. We hypothesised that if facial appearance reflects CNS reliability, then facial attractiveness should be correlated with the variability of behavioural outputs.

Interest in understanding variability *within* individuals is not new (Thouless, 1936; Woodrow, 1932), but it has not been widely acknowledged. Over 50 years ago, Fiske and Rice (1955) conducted

* Corresponding author.
E-mail address: r.ramsey@bangor.ac.uk (R. Ramsey).

http://dx.doi.org/10.1016/j.cognition.2016.10.012
0010-0277/© 2016 Elsevier B.V. All rights reserved.

a systematic review showing that within-person variability – fluctuation in performance across trials or sessions – is not random, but stable, and provides an enduring marker of underlying function. As we describe below, the stable nature of within-person variability and its functional importance have been further supported in the following decades (MacDonald, Nyberg, & Backman, 2006), and is now most frequently assessed by standard deviation in reaction time (RT) across multiple trials (Li, Huxhold, & Schmiedek, 2004). Under this view, rather than reflecting measurement error that should be ignored, consistency of performance is predictive of psychophysiological function (Fiske & Rice, 1955; Thouless, 1936).

Using a variety of RT tasks, evidence has accumulated across cognitive, neurobiological, behavioural and health levels to demonstrate that increased RT variability is associated with reduced functional capacity of the CNS (MacDonald et al., 2006). At a cognitive level, reduced working memory, attention regulation, inhibition, and processing speed have been associated with RT variability (Kofler et al., 2013). At a neurobiological level, higher variability is associated with reduced structural and functional integrity of large-scale brain networks (MacDonald, Li, & Backman, 2009), as well as altered neurotransmitter function (Li & Rieckmann, 2014). In terms of health outcomes, RT variability predicts long-term mental and physical health with healthier individuals showing more consistent performance and those in a diseased state fluctuating more (MacDonald, Hultsch, & Dixon, 2008; MacDonald et al., 2006). Also, as human biology deteriorates in older age, performance on a range of tasks becomes more variable (Li et al., 2004). In sum, cognitive and neurobiological systems are more intact, efficient and healthy in individuals with more consistent performance and compromised in individuals with more varied performance. In addition, reaction time variability and its neural underpinnings have been shown to be heritable, using both quantitative (McLoughlin, Palmer, Rijsdijk, & Makeig, 2014) and molecular genetic approaches (Saville et al., 2014, 2015). In sum, a wealth of evidence from a range of methods supports the importance of processing reliability, as assessed by RT variability, as an important correlate to cognitive, neural, and health-related measures. Consequently, a mate-choice preference for low variability would produce indirect benefits through connection to a high-fitness partner.

Although the cognitive, neural and health correlates of RT variability are becoming clearer, the relationship between RT variability and social signalling remains unknown. Given the large investment of the human species in CNS operation, perceptible cues to CNS function would be expected to be identified and exploited. In particular, we predicted that any visual correlates to CNS reliability should be perceived as attractive. To test whether CNS reliability is visible and attractive, we created composite images from a dataset of 230 individuals who had a headshot photo taken and completed an RT task, which involved raising one of two fingers in response to numerical cues (Fig. 1A). In our first experiment, composite images were made of the 15 individuals from the dataset with the most variable (highest standard deviation of reaction time, SDRT) and least variable (lowest SDRT) latency distributions, for men and women separately. These composite images were shown to a new set of observers who were asked to pick which was more attractive and give an attractiveness rating for each face (Fig. 1B). We measured how frequently low SDRT faces were chosen as more attractive than high SDRT faces, as well as the difference in attractiveness ratings between low and high SDRT faces. If nervous system reliability is signalled through the face, then attractiveness judgments should be associated with low RT variability.

2. Experiment 1

2.1. Method

2.1.1. Participants

58 participants (29 female, M_{age} = 20.3 years, SD = 3.2) took part in the experiment. All participants had normal or corrected-to-normal vision and provided written informed consent prior to data collection. The data reported here were obtained under approval from the Research Ethics and Governance Committee of the School of Psychology at Bangor University. One participant completed the discrimination task but not the ratings task (see below for task details). Thus, 57 participants were included in the analysis of ratings data (28 female, M_{age} = 20.3 years, SD = 3.3).

2.1.2. Stimuli

In total, four composite images of faces were used (see Fig. 1). Based on prior research (Kramer & Ward, 2010; Penton-Voak, Pound, Little, & Perrett, 2006), 15 individual face images were "averaged" using a software package that enables multiple individual faces to be combined into one average face (JPyschomorph; Tiddeman, Burt, & Perrett, 2001). Separately for males and females, the composite images comprised face images from 15 individuals with the highest SDRT and lowest SDRT. SDRT was measured from a sample of 230 participants performing a cognitive control task (for full details, see Butler, Ward, & Ramsey, 2015).

The cognitive control task was developed by Brass et al. (2000) and requires participants to hold down two keys on a computer keyboard and lift one finger in response to a number cue, as quickly and accurately as possible. Simultaneously participants observe a congruent or an incongruent finger movement. Differences between congruent and incongruent conditions were not relevant to the current study and are reported elsewhere (Butler et al., 2015). As such, SDRT is calculated across all 60 trials. Importantly, reaction time variability is a relatively stable construct, which has been shown to have good test-retest and odd-even reliability metrics (Saville et al., 2011). In addition, factor analytic approaches have shown that single factor solutions have normally been adequate to summarise reaction time variability across a number of tasks (Saville et al., 2012; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). Thus, it is likely that the task used here would yield comparable measures of reaction time variability to most other conventional reaction time tasks.

To calculate SDRT scores in order to make the stimuli for the current study, we first excluded participants who were <50% accurate for either condition. We then excluded trials where RT was <100 ms, or >1500 ms, as Ratcliff (1993) showed that this improved power to detect changes in the tau component of the ex-Gaussian distribution, which has been shown to be highly correlated with SDRT (Saville et al., 2011). This led to the exclusion of less than half a percent of the overall number of RTs. Finally, we computed standard deviations for congruent trials and incongruent trials separately and took a mean of these so that each participant had one average SDRT score. Individuals were then ranked according to SDRT. Separate rank orders were produced for males and females in order to generate separate male and female composite images. Face images of the 15 individuals with the lowest SDRTs were then combined into a composite image (low SDRT). The same process was followed using the 15 individuals with the highest SDRTs (high SDRT). The age range of included individuals was narrow for female (low SDRT: 18–27; high SDRT: 18–26) and male composites (low SDRT: 18–26; high SDRT: 18–22) and did not differ between low and high composites (female mean