Neural Networks 24 (2011) 652–664

Contents lists available at ScienceDirect

Neural Networks

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

2011 Special Issue

A computational model of dysfunctional facial encoding in congenital prosopagnosia

Rainer Stollhoff [a,*], Ingo Kennerknecht [c], Tobias Elze [a], Jürgen Jost [a,b]

[a] Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[b] Santa Fe Institute for the Sciences of Complexity, Santa Fe, NM, USA
[c] Institute of Human Genetics, Westfälische Wilhelms-Universität, Münster, Germany

ARTICLE INFO

Keywords:
Infomax
ICA
Independent component analysis
Face recognition
Prosopagnosia
Hereditary prosopagnosia
Congenital prosopagnosia
Faceblind

ABSTRACT

Congenital prosopagnosia is a selective deficit in face identification that is present from birth. Previously, behavioral deficits in face recognition and differences in the neuroanatomical structure and functional activation of face processing areas have been documented mostly in separate studies. Here, we propose a neural network model of congenital prosopagnosia which relates behavioral and neuropsychological studies of prosopagnosia to theoretical models of information processing.

In this study we trained a neural network with two different algorithms to represent face images. First, we introduced a predisposition towards a decreased network connectivity implemented as a temporal independent component analysis (ICA). This predisposition induced a featural representation of faces in terms of isolated face parts. Second, we trained the network for optimal information encoding using spatial ICA, which led to holistic representations of faces. The network model was then tested empirically in an experiment with ten prosopagnosic and twenty age-matched controls. Participants had to discriminate between faces that were changed either according to the prosopagnosic model of featural representation or to the control model of holistic representation. Compared to controls prosopagnosic participants were impaired only in discriminating holistic changes of faces but showed no impairment in detecting featural changes.

In summary, the proposed model presents an empirically testable account of congenital prosopagnosia that links the critical features — a lack of holistic processing at the computational level and a sparse structural connectivity at the implementation level. More generally, our results point to structural differences in the network connectivity as the cause of the face processing deficit in congenital prosopagnosia.

1. Introduction

Faces are a special class of visual stimuli. They are rapidly detected in images and provide a multitude of different information important for social communication such as gaze direction, facial expressions, age, gender, and identity. Under normal conditions of cortical maturation and development, faces are processed in a distributed, hierarchical neural system of face perception (Gauthier, Tarr et al., 2000; Haxby, Hoffman, & Gobbini, 2000; Hoffman & Haxby, 2000; Kanwisher, McDermott, & Chun, 1997). More specifically, facial identity is processed primarily along a ventral occipito-temporal stream with refined processing steps proceeding from the basic analysis of isolated facial features to the structural encoding of holistic, partially view-dependent individual face representations to the establishment of modality-independent personal recognition memories (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005).

Prosopagnosia, colloquially also referred to as "face-blindness", is defined as a profound deficit in the specific task of face identification (Bodamer, 1947). This deficit can be either acquired due to brain damage (see e.g. Mazzucchi & Biber, 1983, for a review of 74 cases), or it is present from birth, i.e. congenital (Behrmann & Avidan, 2005; Grueter et al., 2007; Hasson, Avidan, Deouell, Bentin, & Malach, 2003; Kennerknecht, Pluempe, & Welling, 2008; Kress & Daum, 2003). Congenital prosopagnosia (CP) is highly familial (De Haan, 1999; Duchaine & Nakayama, 2006; Kennerknecht et al., 2006; Kennerknecht, Ho, & Wong, 2008; Kennerknecht, Pluempe et al., 2008; Kennerknecht, Plümpe, Edwards, & Raman, 2007; McConachie, 1976). We therefore coined the term hereditary prosopagnosia (HPA) which can be used synonymously to CP (Kennerknecht et al., 2006). Yet, when initially asked most index subjects are not aware of other impaired family

* Corresponding author.
E-mail address: rainer.stollhoff@mis.mpg.de (R. Stollhoff).

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.03.006

members unless actively interviewed and probed into. Behavioral studies of congenital prosopagnosia have revealed a dissociation between face and object recognition deficits (Duchaine, 2006; Duchaine & Nakayama, 2005; Gauthier, Behrmann, & Tarr, 2004), between face detection and face recognition (Garrido, Duchaine, & Nakayama, 2008), and between the processing of facial identity and facial expressions (Humphreys, Avidan, & Behrmann, 2007), either by testing single aspects in isolation or by conducting a battery of tests with the same participants (Behrmann, Avidan, Marotta, & Kimchi, 2005; Garrido, Furl et al., 2009; Le Grand et al., 2006; Schmalzl, Palermo, & Coltheart, 2008).

The original symptomatic characterization of prosopagnosia by Bodamer (1947) clearly states what prosopagnosia is. But it only includes a vague specification of the processing differences underlying the deficit: "With unimpaired perception of the formal parts of physiognomies, the process of recognition fails" due to an inability to perceive "the structured picture making up an individual, personal whole" (translations taken from Ellis & Florence, 1990). Following up on this characterization of the processing deficits at the computational level we proposed that in CP the failure in integrating information is compensated by a strategy of serially processing informative face parts in isolation (Stollhoff, Jost, Elze, & Kennerknecht, 2010). Such a serial, featural processing can explain the observation of more frequent eye-movements and dispersed gaze behavior in CP (Schmalzl, Palermo, Green, Brunsdon, & Coltheart, 2008; Schwarzer et al., 2007), and an increase in inspection or reaction times (Behrmann et al., 2005; Stollhoff et al., 2010). As an intermediate step between processing faces via isolated face parts or as undifferentiated wholes, holistic encoding (Farah, Wilson, Drain, & Tanaka, 1998), deficits in processing changes in the configuration of features, i.e. the spatial arrangement of face parts, have been documented in cases of acquired prosopagnosia (Barton & Cherkasova, 2005; Barton, Press, Keenan, & O'Connor, 2002; Barton, Zhao, & Keenan, 2003).

So far, prosopagnosia has only been modeled in the acquired case where an existing, functional face recognition system suffers from an externally inflicted damage. The models of acquired prosopagnosia can be roughly divided into two classes: Conceptual models with a focus on neurophysiological correspondence between the functional deficit and the location of the lesion (Breen, Caine, & Coltheart, 2000; Ellis & Lewis, 2001; Fox, Iaria, & Barton, 2008), and abstract, computational models with a focus on task differences in the required recognition or recall accuracy (Burton, Bruce, & Hancock, 1999; Farah, O'Reilly, & Vecera, 1993; Pessa, Bandinelli, & Penna, 1999; Virasoro, 1988, 1989; Zifan, Gharibzadeh, & Moradi, 2007). In both classes of models a fully functional, mature system is degraded, e.g. by removing nodes or clipping connections. Evaluation of the model is then based on comparing the properties of the network before and after degradation either descriptively, analytically, or numerically using simulation studies. Here, a neural network model of congenital prosopagnosia (CP) is derived from formal considerations, implemented in an artificial neural network model of facial encoding, and tested empirically in experiments with prosopagnosic and control participants. The main accomplishment of the model is to provide a direct, testable link between the critical features of congenital prosopagnosia as the lack of holistic processing at the computational level and a reduced structural connectivity of face processing areas at the level of neuronal implementation.

1.1. Neuroanatomy of face processing

Under normal conditions the brain develops a specialized neural system for face recognition (de Haan, Humphreys, & Johnson, 2002; Polk, Park, Smith, & Park, 2007; Scherf, Behrmann, Humphreys, & Luna, 2007) which has been further differentiated into spatially segregated functional processing modules (Gauthier, Curby, Skudlarski, & Epstein, 2005; Gauthier, Tarr et al., 2000; Grill-Spector, Knouf & Kanwisher, 2004; Haxby et al., 2000; Hoffman & Haxby, 2000; Kanwisher et al., 1997; Kawashima et al., 2000). Irrespective of the exact developmental processes underlying the functional specialization, damage inflicted to a specific cortical region can therefore lead to restricted deficits conditional on the interconnectedness and interdependence of the distributed processing (Damasio, Tranel, & Damasio, 1990; De Renzi, Faglioni, Grossi, & Nichelli, 1991; De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994; Fox et al., 2008). More specifically, the behavioral heterogeneity in acquired prosopagnosia can largely be explained by differences in the extent and location of the brain damage causing the deficit (Damasio et al., 1990; De Renzi et al., 1991, 1994; Fox et al., 2008).

In contrast, functional imaging studies of CP have so far found no unequivocal evidence for activation differences in this region; neither using classical localizer paradigms (Avidan, Hasson, Malach, & Behrmann, 2005; Hasson et al., 2003) nor adaptation paradigms (Avidan & Behrmann, 2009; Avidan et al., 2005). First indications of structural neuroanatomical differences point to a volumetric reduction of the anterior fusiform gyrus (Behrmann, Avidan, Gao, & Black, 2007) and the anterior inferior temporal lobe (Garrido, Furl et al., 2009), regions involved in more associative and mnestic aspects of face recognition (Haxby et al., 2000). Analysis of a large group of CP participants, revealed diminished gray matter density in the lingual gyrus bilaterally, the right middle temporal gyrus and the dorsolateral prefrontal cortex (Dinkelacker, Grueter, Klaver, & Grueter, in press). In a diffusion tensor imaging study, Thomas et al. (2009) reported a reduced structural connectivity in the ventral occipito-temporal white matter tracts, presumably involved in more apperceptive aspects of face recognition. In our model, these observations provide the rationale for implementing the structural differences in CP as a predisposition towards a reduced network connectivity between regions involved in the structural encoding of facial information.

1.2. Modeling principles

Possible alterations in the process of structural encoding of face images will be studied in the framework of single-layer feedforward networks. The input units of the network register the stimulus which is then encoded into a sensory description based on the activation of the output units. In a single-layer feedforward network, the input stimulus $X = (X_1, \ldots, X_D)$ is mapped to the output activation $S = (S_1, \ldots, S_n)$ by

$$S_j = h\left(\sum_{d=1}^{D} w_{jd} X_d\right),$$

where $w_{jd} \in \mathbf{R}$ is the weight associated with the connection from the dth input unit, X_d, to the jth output unit, S_j, and h is the activation function, in matrix notation: $S = h(\mathbf{W}X)$. The network thus translates an observable input vector or stimulus X into an internal representation S, which can be used for further processing. While in this formulation the projection of input to output units involves only feedforward computations, training of the network, e.g. weight adaptation, often draws on information that is not available locally, e.g. the activations of the other output units.

For our models of functional and dysfunctional facial encoding we trained single-layer feedforward networks to represent a set of frontal face images under two different constraints (see Fig. 1). On the one hand, we introduced a constraint on the sparseness of the output unit activation which was implemented at the algorithmic