Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

The Early Time Course of Compensatory Face Processing in Congenital Prosopagnosia

by

Rainer Stollhoff, Jürgen Jost, Tobias Elze, and Ingo Kennerknecht

Preprint no.: 8 2010

The Early Time Course of Compensatory Face Processing in Congenital Prosopagnosia

Rainer Stollhoff[1,*], Jürgen Jost[1,2], Tobias Elze[1], Ingo Kennerknecht[3]

1 Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

2 Santa Fe Institute, Santa Fe, NM, USA

3 Institute of Human Genetics, Westfälische Wilhelms-Universität, Münster, Germany

∗ **E-mail: rainer.stollhoff@mis.mpg.de**

Abstract

Background

Prosopagnosia is a selective deficit in facial identification which can be either acquired, (e.g. after brain damage), or present from birth (congenital). Previously, the face recognition deficit in prosopagnosia has been characterized by worse accuracy, longer reaction times, more dispersed gaze behavior and a strong reliance on featural processing. Here, we portray compensatory processing in congenital prosopagnosia as a serial inspection of diagnostic features.

Methods/Principal Findings

We investigated performance differences in different face and shoe identification tasks between a group of 16 participants with congenital prosopagnosia and a group of 36 age-matched controls. Given enough training and unlimited stimulus presentation prosopagnosics could achieve normal face identification performance albeit at the expense of longer reaction times. This increase in reaction times can be accounted for by an equally-sized increase in stimulus presentation times needed by prosopagnosics to perform at the same level as controls. Reversely, as the inspection time of face stimuli was limited by tachystocopic presentation, prosopagnosics only showed worse performance but differences in reaction time vanished. Thus, prosopagnosics inspect faces longer but require a normal amount of time to make and execute the decision. In a further experiment, rotation of face stimuli away from the learned frontal viewed decreased recognition performance for both groups similarly. All group differences in performance, reaction or presentation times, and the influence of rotation were selective to face stimuli and didn't extend to shoes.

Conclusions/Significance

Our study provides a characterization of congenital prosopagnosia in terms of early processing differences. More specifically, compensatory processing in congenital prosopagnosia requires an inspection of faces that is sufficiently long for sequential focusing on diagnostic features. Our characterization of dysfunctional processing in prosopagnosia further emphasizes fast and holistic information encoding as the defining characteristic of normal face processing.

Introduction

Prosopagnosia, colloquially also referred to as "face-blindness", was first defined by Bodamer as a selective deficit in the specific task of face identification [1], although the deficit has been reported previously in conjunction with more general object recognition deficits [2–4]. Since then, prosopagnosia has mostly been observed in cases of acquired prosopagnosia, where the deficit was caused by neurological damage following e.g. intoxication, head injury or encephalopathy [1,5–9].

Recently, more and more cases of prosopagnosia have been reported where the deficit was not acquired due to an accident, but presumably present from birth, i.e. congenital [9–17]. In contrast to the rare acquired form, the congenital form is among the most common anomalies in humans with a prevalence of 2.5 % [16], and is almost always hereditary [17–19]. Notwithstanding ongoing discussions on the nature of this congenital form, here we will continue to refer to all cases of prosopagnosia without any exogenous cause as cases of congenital prosopagnosia (CP), without explicitly addressing the question of heritability or developmental influences.

The face recognition deficit in CP can be as profound as in the acquired form [9] and equally selective such that only facial identification is impaired while all other aspects of face and object recognition remain intact [20]. However, recent reports on deficits in the processing of biological motion [21] and impaired visual mental imagery [22] question the selectivity. Overall, cases of CP often display heterogeneous symptoms [23] which has so far prevented a stringent categorization of congenital prosopagnosia according to phenotypical symptoms.

Previously, face processing in CP has been characterized by a reliance on featural processing [15, 24] as compared to holistic processing in control participants [25]. Furthermore, while controls tend to fixate mostly on the eye regions to discriminate between faces [26, 27] and perform worse if during learning fixations are restricted to other regions [28], CP participants display more dispersed fixation patterns [29,30]. Also, participants with CP can achieve close to normal face recognition performance in standardized tests [31], however they often show longer reaction times [15] and a more pronounced deficit under tachystocopic presentation [32,33].

Here, we investigate the face recognition deficit in CP based on the following qualitative model of facial information processing. We propose that normally faces are encoded by a fast and parallel integration of information that is distributed over the whole face (holistic); in order to minimize the influence of image translations, humans adopt a stereotypical fixation pattern. During recognition the perceived snapshot is matched in parallel against stored snapshots. Face recognition is thus a mostly perceptual and associative process, involving very limited conscious effort. Thus in an ideal observer model, facial encoding can be thought of as recording an informative snapshot of a full face image, where informative regions (e.g. the eyes) are encoded with a higher precision (e.g. retinal and cortical "resolution") than non-informative ones. Informativeness of regions is learned over repeated exposures to faces and depends on the population exposed to (cf. "other-race-effect" [34]) and only to a lesser degree on the individual face stimulus encoded. During encoding the overall precision of the stored snapshot increases over time as perceptual noise is discounted. During recognition a stimulus is matched in parallel to stored snapshots and the precision is limited not only by perceptual noise but moreover by the precision of the previously stored snapshots.

In contrast, in CP facial encoding is mostly a deliberate process of extracting local (featural) information in a series of fixations or attentional shifts. As the uniqueness of isolated features,

and therefore their informational content, differs between individual faces, so does the series of fixations employed to extract the information. In contrast to normal face processing, this compensatory processing is contingent on cognitive strategies of problem solving, which is in agreement with a face-specific increase in the BOLD response in frontal areas in congenital prosopagnosia [35]. Thus, in an ideal CP observer model, during the initial encoding of the stimulus a face image is scanned for informative regions and if given sufficient time a unique, optimally informative series of fixations are developed for every single face. During recognition the fixation sequence is iteratively adapted according to prior evidence as well as the information extracted by prior fixations.

A central focus of this work are temporal dynamics of face recognition and the influence of stimulus transformations. The proposed model of processing differences between controls and CPs makes the following predictions: First, if given sufficient training and inspection times, CP participants might be able to achieve the same performance as controls [15, 31] presumably by serial matching of isolated features [29]. Second, for any given fixed inspection time, on average CP participants will extract less information and thereby perform worse [32, 33]. Third, for controls, limiting inspection time during initial encoding has a stronger influence than limiting inspection time during recognition as matching precision is limited by the precision of stored snapshots. Fourth, we hypothesized that in both cases processing relies mostly on appearance-based (i.e. pictorial) information that doesn't generalize well across rotation in depth [36]. Thus, the deficit in CP is not related to a dysfunctional generalization and on average the influence of stimulus transformation shouldn't differ between control and CP participants.

In order to test the proposed model and better characterize the behavioral symptoms in congenital prosopagnosia, we conducted a series of experiments, each testing different aspects of face and object recognition, with a total of 16 CP and 32 control participants. The setup of the experiments closely parallels those of other tests (e.g. Cambridge Face Recognition Test [37]), but was restricted to setups which have a direct analogue in real-life situations avoiding unrealistic conditions, e.g. scrambling or inverting a face. Most of the experiments were conducted with human faces as well as non-face stimuli drawn from the same object category (Nike^{TM}sneakers resp.) to investigate face-specificity of the deficit [33].

In the first two experiments, a standard setting was used to test recognition of frontal images of faces (experiment 1) and shoes (experiment 2). Participants were familiarized with four individual target stimuli and later on had to identify the targets amongst a group of distractor stimuli in a two-alternative forced choice (2-AFC) paradigm . We specifically investigated whether longer reaction times can be attributed to longer inspection of the images or a longer decisional component. First, we measured participants' reaction times under the condition of unlimited presentation. Second, we used an adaptive sampling strategy to estimate the presentation time at which a participant performs with an accuracy of 80%. Third, we contrasted individuals' reaction times with their 80%-correct presentation times.

In experiment 3 (faces) and experiment 4 (shoes) we investigated each participants' ability to generalize stimulus recognition across rotation in depth. While recognition of stimuli taken under identical viewing conditions can be solved by image matching, rotation in depth which occurs frequently under natural viewing conditions at least diminishes the applicability of similar compensatory strategies. Previously, it was shown that for faces learned in a frontal view, recognition performance decreases monotonically when tested with images rotated around the

vertical axis [36]. Here, we assessed whether participants with CP show a similar or more pronounced decrease in performance. In order to isolate the influence of rotation and avoid statistical ceiling (or bottom) effects in the performance, images were displayed for different durations estimated according to individual performance in experiments 1 and 2 respectively.

The setup of experiments 1 and 2 entailed the presentation of stimuli for different durations that were individually determined for each CP participant and their respective matched controls. In experiments 5 and 6, four fixed presentation times were used, identical for all participants and chosen to separate between the times needed for preparation and execution of one or multiple saccades. In addition, we investigated whether there are differences in the effect of tachystocopic presentation depending on whether they are applied during the encoding, i.e. learning, of a novel face (experiment 5) or during the decoding, i.e. recognition, of a previously learned face (experiment 6).

Differences in the age of participants as well as observations of cognitive heterogeneity in CP participants [23] prompted us to resort to a threefold statistical analysis: First we compared raw test results. Second, we fitted generalised linear mixed models to account for age related changes and to test for group differences in the influence of experimental parameters on participants' performance. Third, for each participant we calculated individual residuals as the difference between actual outcomes and the outcome that would be expected based on average control performance.

Materials and Methods

The experiments were conducted at different times and locations. Experiments with CP participants took place at the Institut für Humangenetik, Westfälische-Wilhelms-Universität, Münster, experiments with control participants took place at the Max Planck Institute for Mathematics in the Sciences, Leipzig. Experiments 1-4 were conducted at the end of 2006, experiments 5 and 6 a year later at the end of 2007.

Participants

In total we tested 16 CP participants and 36 age matched controls (two for each CP participant). Participants age (at first testing, i.e. end of 2006) varied between 20 and 68 years for the CPs (mean: 37.9, sd: 17.5) and for the controls (mean: 36.6, sd: 15.8).

Except for one CP participant (MB) all CP and control participants reported normal or corrected to normal vision. MB has a strabismus convergens, on which she was operated on three times during childhood. However, she still reported on perceiving diplopic images and difficulties with stereopsy.

Ethics Statement All CP and control participants provided written informed consent before participation. The study was approved by the ethical committee of the University of Muenster, Germany, protocol No 3XKenn2.

Participants with Congenital Prosopagnosia

All of the 16 CP participants were diagnosed using a semi-structured interview [16, 17, 19, 38], which includes questions on everyday-problems with face and object recognition, mental imagery and avoidance strategies.

CP participants - and accordingly control participants - fall into two different age groups: one consisting of 8 younger CP participants, most of them students, aged between 21 and 26 years, the other consisting of 8 older CP participants, aged between 41 and 68.

- The younger group (born after 1980) consisted of five students of medicine, which were detected by a screening study which was conducted by means of a questionnaire (see [39] for a detailed description). Students who reported suspicious behavioral deficits were then invited for the semi-structured interview. In addition, three of the younger participants established contact after having been informed about prosopagnosia via public media or personal contact.
- The older group (age 40 or older) is composed only of people who initiated contact themselves.

See table 1 for a short overview of CP participants.

Three CP participants (JM, HW, LL) only participated in experiments 1-4. For one CP participant (MB) online estimation of the 80%-correct presentation time in experiment 1 failed due to technical problems. We therefore excluded MB and both matched controls from all analysis of experiments 1 and 3 which involved limited presentation times.

Most CP participants had intact basic-level object recognition abilities as assessed by a total of seven subtests chosen from the Birmingham Object Recognition Battery [40] and the Visual Object and Space Perception Battery [41]: Ten CP participants scored in the normal range in all tests, only two (HW,HB) had difficulties in more than one of the seven subtests, and none had difficulties in more than three subtests.

Control Participants

In experiments 1-4 a total of 32 control participants were selected to match the age of CP participants and also gender in most cases. Younger control participants were mostly students, similar to CP participants, while the older control participants showed a similar variety in profession as the older CP participants. In experiments 5 and 6 a total of 24 age and mostly gender matched controls participated, 21 of whom also had participated in experiments 1-4.

Experiments 1 to 4

Stimuli

The face stimuli were obtained from the publicly available Face Database of the MPI for Biological Cybernetics (see [36] for details on the database creation) which contains snapshots of 3D-scans of 200 heads (without hair), taken at seven rotations (frontal view and 3 rotations in each direction of 30°, 60°and 90°). Snapshots were used as distractor stimuli. Target face stimuli were generated using the four individual full head models in the Face Database

(two male and two female heads). Snapshots of the full head models under the same rotations (30°, 60°and 90°) were generated using Blender free open source 3D content creation suite (http://www.blender.org, open-source). All snapshots are 8-bit color images of 256x256 pixels.

The shoe stimuli were obtained as snapshots of different sneakers obtained from http://nikeid.nike.com. A total of 53 distractor shoes and 4 target shoes were used, all under the available three different rotation conditions (oblique, side and top view).

A randomly chosen subset of 16 distractor objects was used during the learning blocks, and the remaining distractor items were used in the testing blocks. This split ensured that participants learned to recognize targets and not distractors. All matched controls had exactly the same experimental setup (choice of distractors objects during learning and testing as well as presentation order) as their respective CP participants.

Design

Experiments 1 to 4 all started with a simultaneous presentation of all 4 target images in frontal view for unlimited duration, which was then followed by a feedback training. Training consisted of 16 trials, 8 of which showed the targets (2 times each) and 8 showed a distractor. The participant had to respond by clicking the left mouse button for a target and the right mouse button for a distractor (two alternative forced choice - 2-AFC). In the training feedback (correct/false) was given after the response. Presentation and feedback training were repeated until participants made at most a single error during the 16 training trials. The presentation order of target and distractor images was randomized in each training block, but exactly the same for a CP participant and his/her two matched controls.

After successful completion of the training the test started (see figure 1 A for a schematic depiction of the experimental design). In the midst of each test another round of presentation and feedback training was administered.

Testing in experiment 1 and 2 The first experiments for each stimulus class (faces or shoes) tested recognition of the frontal view of target images under varying presentation times (PT) in a two alternative forced choice task. The presentation times were chosen according to the accelerated stochastic approximation method [42, 43] with a threshold at 80% correct, see below. The algorithm increased presentation times, whenever a mistake was made and decreased presentation time after a correct answer. Trials were grouped into blocks of 8 such that in each block every target appeared exactly once. Presentation order of target and distractor image was randomized, but equal for a CP participant and his/her two matched controls.

Testing in experiments 3 and 4 For each stimulus class, we tested recognition of targets in the frontal view and under rotations in a two alternative forced choice task (for faces: 3 in-depth rotations of 30°, 60°and 90°in each direction; for shoes:side view and top view). For faces the test contained two testing blocks of 56 images each (7 rotation conditions with 4 targets and 4 distractors each); for shoes it contained four testing blocks of 24 images each (3 rotation conditions with 4 targets and 4 distractors each). Presentation order of target and distractor stimuli, and rotation angles was randomized in each block, but equal for a CP participant and his/her two matched controls. The presentation time was set to a fixed value chosen for each

CP participant individually as an estimate of the time that he/she would need to give correct answers 90% of the time if tested with frontal face images (see below for details of the estimation process). This presentation time was also used for both matched controls of the CP participant.

Experiments 5 and 6

Stimuli

All stimuli were generated with the assistance of the Recognition and Categorization Group in the Department Bülthoff at the Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany. Face images were obtained by rendering from a total of 96 full 3D head models. The acquisition of 3D-scans and the generation of the head models is described in [36].

For each test we selected 48 individuals, 24 male and 24 female faces. For each individual face we rendered 5 face images, differing in rotation and illumination. One reference image (target stimulus) was taken in frontal view with ambient illumination (rgb= 0.5 0.5 0.5) and an additional white illumination (rgb=0.7, 0.7, 0.7) from a direction in front, above and to the right of the face (horizontal rotation=50 °, vertical rotation=50 °). Four test images were taken under slight horizontal and vertical rotations of $(\pm 5, \pm 5)$ degrees. In all test images the position of the white illumination source was changed to come from in front, below and to the left of the face (horizontal rotation=-50 °, vertical rotation=-20 °).

All reference images were standardized to the same rectangular area (i.e. width x height) of the facial image at roughly 25000 square pixels. Resulting images had widths between 122 and 138, and heights between 181 and 204 pixels. Size variations (standard deviation divided by mean) in width and height were slightly smaller compared to human anthropometric measurements [44]. Resulting images were placed upon a black background such that each face was in the center of a 140x210 pixels image.

Test images were not standardized as the standardization in reference images already discounted all size differences with respect to a scaling of the whole image.

Experimental Design

In experiment 5, a target stimulus was first presented for a short duration of 50, 150, 450 or 750 ms, and after a blank of 500 ms, the target stimulus had to be recognized among a total of four face stimuli (see figure 2 A). The response was indicated by pressing one of the four arrow keys corresponding to the position of the stimuli (left, right, up or down). During the recognition the four face stimuli were presented until a response was made. The position of the target stimulus in the test display was randomized. All combinations of 48 target stimuli and four presentation times were tested exactly once per participant, yielding a total of 192 trials. As each stimulus was used both as target in four trials and as distractor in 12 trials, the order of the trials in which each stimulus was presented as target or as distractor was counterbalanced across presentation times to exclude influences of familiarity on the estimation of presentation time effects.

In experiment 6, four target stimuli were initially learned over three rounds of unlimited presentation and subsequent feedback training, and afterwards the influence of variable presentation time on recognition performance was tested (see figure 2 D).

In each training trial either one of the four target stimuli or one of four different distractors was shown and participants had to indicate, whether the presented stimulus was one of the four target stimuli. Each training round included 8 training trials, where each target and each distractor were presented exactly once in a randomized order.

In each test trial participants had to indicate whether the presented stimulus was one of the four target stimuli learned in the previous training. Each test round consisted of 32 test trials: 16 target presentations and 16 distractor presentations. In every round, each target was presented once for each duration of either 50, 150, 450, or 750 ms. The four distractor stimuli used in the training and an additional four new distractor stimuli were displayed two times each. In these 16 distractor presentations, each of the four presentation durations was chosen four times, but no distractor stimulus was shown at the same duration twice.

The full cycle of presentation, feedback training and test was repeated 12 times such that each of the 48 face stimuli was exactly once among the target stimuli, once among the distractors already present during the training and once among the additional distractors introduced in the test round. Presentation order was counterbalanced across stimuli, groups and presentation time.

In both tests - but not in the feedback training in experiment 6 - test and reference images differed in size, rotation and illumination (see above).

The presentation times of 50, 150, 450 or 750 ms were chosen to separate between different types of eye movements. As the experiment should be applicable to participants of different age, we relied on studies explicitly addressing age differences in the times needed to perform saccadic eye-movements [45]. However, it is possible that true saccade times found in face processing deviate [27].

Presentation

In experiments 1-4 images were displayed on either an IIYAMA Vision Master Pro514 monitor (22', at 200 Hz) or an IIYAMA Vision Master 506 (21' at 170 Hz) (random assignment, identical for each CP and his/her matched controls) both running at a 800 x 600 resolution with a screen area of approximately 400 mm x 300 mm. Participants were initially seated at a distance of 1m and stimuli covered on average 140 pixels x 210 pixels, i.e. 70 mm x 110 mm, equal to a visual angle of 4.0°horizontally and 6.3°vertically at a distance of 1 m.

In experiments 5 and 6 presentation was always on the IIYAMA Vision Master Pro514 monitor (22', at 200 Hz) used previously with a resolution of 800 x 600 and images subtended 130 pixels x 190 pixels, i.e. 65 mm x 85 mm or 3.5°x 4.3°at the initial seating distance of 1 m.

All experiments were run using the open-source flashdot experimental psychophysics presentation software ([46], available at `http://www.flashdot.info`), which allows a high temporal precision of the presentation. Presentation duration was actually measured in frames, durations are reported in ms for convenience and to enable comparisons between results obtained at difference monitors with different frame rates. To convert between frames and ms, we simply multiplied the number of frames by the inverse frame rate, which can deviate from the actual presentation times for very small durations [47].

Statistical Analysis

If not noted otherwise, all data analysis and statistical testing was done in the statistical programming language R [48].

Estimation of Presentation Times

In experiments 1 and 2, accelerated stochastic approximation [42] was used as an online estimation method to obtain a presentation time at which participants would make correct responses in 80% of the trials ($\mathrm{PT}_{80\%}$). A fraction of 80% correct answers was chosen to avoid ceiling and/or bottom effects and achieve efficient sampling [49, 50]. Given a sequence of at least two answers $(Y_1, \ldots, Y_n)$, where Y_k is 1 for a correct answer and 0 for an incorrect answer and an initial presentation time (PT_0), presentation times in the $n-th$ trial (PT_n) were adjusted according to

$$PT_{n+1} = PT_n - \frac{500\,\mathrm{ms}}{2 + \#\left\{Y_k \neq Y_{k-1}\right\}_{k<n}} (Y_n - .8), \tag{1}$$

where the denominator contains a counter of the number of shifts in the answer from correct to incorrect or vice-versa.

As the participants completed the tasks on two different monitors with frame rates of 170 Hz and 200 Hz respectively (see below) the actual presentation time sequences had stepsizes of 5 ms and $1/170 \approx 6$ ms. The mean of the presentation times of the last 16 trials was taken as an estimate of $\mathrm{PT}_{80\%}$.

The data obtained in experiments 1 (or 2 resp.), i.e. all pairs $(Y_k, PT_k)_{k=1}^{N}$, was used to estimate a presentation time for each participant at which he/she would achieve 90% correct answers ($\mathrm{PT}_{90\%}$). In a pilot study using an 80% correct time we observed bottom effects in performance on rotated images, and thus increased the threshold to 90%.

The software package psignifit (http://www.bootstrap-software.org/psignifit/) was used to fit a modified logistic regression model as an estimate of participants psychometric function relating the response to the presentation time [49]:

$$Pr(Y = 1|PT; \alpha, \beta, \lambda, \gamma) = \gamma + (1 - \gamma - \lambda) \frac{1}{1 + exp\left(-\frac{(PT-\alpha)}{\beta}\right)}. \tag{2}$$

Here, we used as guess rate $\gamma = 0.5$, an upper bound on the lapse rate $\lambda \in [0, 0.05]$ and restricted the shape of the logistic regression function by requiring a positive intercept $\alpha \in [0, 500]$ and slope $\beta \in [0, 500]$. The $\mathrm{PT}_{90\%}$ obtained by inverting the fitted logistic regression model in equation (2) to the data gathered in experiment 1 (or 2 resp.) was then used as a presentation time in experiment 3 (or 4 resp.). Each CP participant as well as the two matched controls were shown images at the $\mathrm{PT}_{90\%}$ estimated for the CP participant.

In an a-posteriori check, the estimates obtained for the 90%-correct presentation time appeared appropriate, as for presentations of frontal images using these $\mathrm{PT}_{90\%}$ estimates the median CP recognition of frontal views (93% for faces, 95% for shoes) was only slightly above the target rate (see figure 1 E).

Generalized Linear Mixed Models

To assess differences in the influence of experimental variables between the control and the CP group, generalized linear mixed models (GLMMs) were used (see e.g. [51] for an introduction to GLMMs). First, a nullmodel that always included fixed effects for age and all experimental variables (e.g. presentation time) as well as random effects allowing for individual variation in the mean and also in the influence of experimental variables was fitted. Based on this nullmodel, alternative, nested models were constructed by subsequently adding group differences in the influence of fixed effects, i.e. firstly a mean difference between the groups (main effect), secondly an interaction of group and experimental variables (first-order effects), ... (see below for details). Comparison of nested models was based on differences in the log-likelihood of the models, i.e. a likelihood ratio test (LR-test).

In addition to model based comparisons, we calculated residuals for each participant using a modified cross-validation approach. The obtained residuals measure the deviation of CP participants performance from the expected performance of a hypothetical control participant with identical individual characteristics (e.g. age). First, parameters of the nullmodel were estimated using only control participants' observations, and the resulting parameter estimates were used to calculate residuals for the observations of CP participants. Second, for control participants, an approach similar to a leave-one out cross-validation was applied: For each control participant, parameters of the nullmodel were estimated using the observations of all control participants except the latter one, and the estimates were then used to calculate residuals for this control participant. Residuals were averaged across observations into a single number for each individual participant. For group comparisons of these average residuals the non-parametric Wilcoxon rank sum test, also referred to as Mann-Whitney test, was used since CP participants' residuals were in most cases not Gaussian-distributed.

Fitting of generalized linear mixed models (GLMMs) was done using the R packages `lme4` [52] and `MCMCglmm` [53]. The algorithms used in `lme4` as well as the model based comparisons conducted here, are described by the main contributor to the `lme4` package in more detail in [51]. To test for significant differences likelihood ratio tests were performed where we assumed a χ^2 distribution of the test statistics with degrees of freedom equal to the difference in the number of parameters. In testing significance of fixed effects in mixed models, the χ^2 approximation tends to produce to p-values that are too small [51]. Hence, if the selected model included interaction effects, the model was again fit with `MCMCglmm` to obtain Bayesian maximum posterior estimates ($\hat{\beta}$))and highest posterior density intervals with 95% support ($\text{HPDI}_{95\%}$) for parameter estimates of interaction effects [54]. As prior distributions for the Bayesian model fitting we used a multivariate normal distribution with zero mean and a diagonal covariance matrix with large variances ($\sigma = 10^{10}$) for fixed effects and an inverse Wishart distribution with degrees of freedom equal to one and the inverse scale equal to the unconditional variance of the response variable.

Experiments 1 and 2 To analyze group differences in reaction times in experiments 1 and 2, a nullmodel including age and training block number as fixed effects was fitted with individual random effects on mean reaction time as well as on the influence of block number. The inverse of the reaction time was taken as the dependent variable to improve model fit (see below). Comparison of group differences in the presentation time needed for 80% correct recognition

($PT_{80\%}$) was based on a linear model. The nullmodel only included age as a fixed effect and $PT_{80\%}$was log-transformed. To assess, whether for faces the group differences found in reaction times can be explained by group differences in presentation times, again nullmodels were constructed for both dependent variables but this time without a prior transformation, and the respective residuals, as well as the difference in residuals was calculated.

Experiments 3 and 4 Differences in the influence of rotation and presentation time on control and CP participants‘ performance in experiments 3 and 4 were tested based on binomial GLMM nullmodels with logit-link including with age, rotation angle (nominal scale according to the absolute values, i.e. 0°, 30°, 60°, and 90°) and presentation time (log-transformed and mean-centered) as fixed effects and participant identity as random effect. Interactions were tested both for rotation and presentation time combined (full interaction model), as well as specifically for group differences in the influence of presentation time (PT interaction model) to account for a priori known differences in the experimental setup (see above).

Experiments 5 and 6 Analysis of recognition rates in experiments 5 and 6 was based on a binomial GLMM nullmodel with logit-link including age and the logarithm of presentation time (during training or during recognition resp.) as fixed effects and participants identity as random effect. Reaction times were transformed by taking the inverse (see below) and then analysed using a LMM nullmodel with the same fixed and random effects.

Prior Adjustments of Reaction Time Observations

In the results of experiments 1 (and 2), we only analyzed reaction times observed during feedback learning trials occurring in experiments 1 and 3 (2 and 4) excluding the very first feedback training in experiment 1 (2 resp.). In addition, we restricted analysis to trials where a correct answer was given (5349 out of 5424 observations). A fixed cutpoint for observations of reaction time was used dismissing observations with reaction times above 2000 ms and below 500 ms to reduce the number of outliers [55]. A total of 5080 observations of reaction time was analyzed, 2908 observations for face stimuli in experiment 1 and 2172 observations for shoe stimuli in experiment 2.

In experiment 5, only reaction times between 500 ms and 8000 ms (corresponding to the 0.01% and 95% quantiles) were included in the analysis.

In experiment 6, only reaction times between 500 ms and 4000 ms (corresponding to the 0.05% and 96% quantiles) were included in the analysis. Note that participants could respond only after the end of the stimulus presentation.

Prior to any analysis using model based comparisons, reaction times were transformed by taking the inverse to improve applicability of linear models [55].

Results

Experiments 1 and 2

For the face stimuli used in experiment 1, CP participants made more mistakes than controls during the initial feedback training. Also, reaction times - measured during subsequent feedback trainings - as well as the presentation times individual participants needed to achieve 80% correct recognition rates were larger among CP participants. Difference in reaction times are of the same magnitude as difference in presentation times, which suggest that increased reaction times for CP participants are due to a longer inspection of the stimulus as opposed to a longer time to reach a decision. For the shoe stimuli used in experiment 2, control and CP participants' performance, reaction and presentation times didn't differ significantly.

Performance during feedback learning On average participants with congenital prosopagnosia made more mistakes and needed more training than controls during the very first learning phase prior to testing in the face recognition test (Wilcoxon rank sum test, $W = 87$, $n_0 = 32$, $n_{CP} = 16$, $p < 0.001$), but not in the shoe recognition test ($W = 226$, $n_0 = 32$, $n_{CP} = 16$, $p = 0.47$).

Reaction times On average CP participants had longer reaction times than controls for faces (LR-test of main effect, $\delta(\updownarrow)= 13.00$, $df = 1$, $p = 0.001$, $\hat{\beta} = -1.7 \cdot 10^{-4}$, $\text{HPDI}_{95\%} = [-2.6, -0.7] \cdot 10^{-4}$), but not for shoes ($D = 0.00$, $df = 1$, $p = 0.95$). In contrast to group differences in the mean, in both cases the influence of training block number on reaction time didn't differ between groups (LR-test of first-order against main effect model, $D = 0.04$, $df = 1$, $p = 0.85$ for faces, and $D = 0.20$, $df = 1$, $p = 0.66$ for shoes).

A comparison of residuals revealed that participants with CP had longer reaction times than expected (larger residuals of -1/(reaction time)) for faces ((Wilcoxon rank sum, $W = 100$, $n_0 = 32$, $n_{CP} = 16$, $p < .001$ one-sided) but not for shoes ($W = 244$, $n_0 = 32$, $n_{CP} = 16$, $p = .40$ one-sided). The increase in reaction time was selective for faces for most CP participants (see figure 1 B) 13 out of 16 CP participants had higher residuals in the face task compared to the shoe task.

Presentation times All control participants only needed very short presentation times to perform at an 80% correct level in both the face (individual $\text{PT}_{80\%}$s range between 6 ms and 47 ms) and shoe recognition task (between 9 ms and 102 ms) which is well below the time needed for controlled eye movements (i.e. less than 200 ms). In contrast, several participants with congenital prosopagnosia required far longer presentation times to accurately recognize faces (21 ms to 766 ms) and/or shoes (7 ms to 462 ms).

Comparing both groups, $\text{PT}_{80\%}$s were larger for CP than control participants for faces (medians of 34.2 ms and 20.9 ms respectively; Wilcoxon rank sum test, $W = 94.5$, $n_0 = 32$, $n_{CP} = 15$, $p < 0.001$ one-sided) but not for shoes (medians of 20 ms and 17.3 ms respectively; $W = 235$, $n_0 = 32$, $n_{CP} = 16$, $p = 0.33$ one-sided). Analogously, using model based comparisons revealed a significant group difference in the log-transformed $\text{PT}_{80\%}$for faces (LR-test, $D = 24.02$, $df = 1$, $p < 0.001$) but not for shoes ($D = 2.52$, $df = 1$, $p = 0.11$).

Compared to the control group CP participants had larger residuals in the log-transformed $PT_{80\%}$- needed longer presentation times than expected - for the face task ($W = 67$, $n_0 = 30$, $n_{CP} = 12$, $p < 0.001$ one-sided) but not for the shoe task ($W = 233$, $n_0 = 32$, $n_{CP} = 16$, $p = 0.31$ one-sided).

Comparison of reaction times and presentation times While CP participants had significantly higher residuals for reaction times ($W = 63$, $n_0 = 30$, $n_{CP} = 15$, $p < 0.001$) recognizing faces in experiment 2, there is no significant group difference left after we subtracted participants' $PT_{80\%}$residuals ($W = 201$, $n_0 = 30$, $n_{CP} = 15$, $p = 0.29$) and thereby accounted for differences in the time participants need to inspect a stimulus (see figure 1 D). Thus, while CP participants needed to inspect face images longer than controls, the time taken to make a decision didn't seem to differ.

Experiments 3 and 4

CP participants made more mistakes in recognizing rotated face images than controls. However, the difference in performance did not change across the rotation angles tested. For shoe stimuli, no significant group difference were observed.

CPs performed significantly worse than controls for faces across all rotation angles (Wilcoxon rank sum tests, $n_0 = 30$, $n_{CP} = 15$, all $p < .05$ one-sided) but there was no significant difference for shoes in any rotation condition ($n_0 = 32$, $n_{CP} = 16$, all $p > 0.3$ one-sided), see figure 1 E.

Model based comparisons revealed a significant main effect for faces (LR-test of main effect, $D = 22.15$, $df = 1$, $p < 0.001$), but not for shoes ($D = 0.03$, $df = 1$, $p = 0.87$). In contrast, for both types of stimuli the influence of presentation times on recognition differed between the two groups (faces: LR-test of PT interaction model against main effect model, $D = 7.84$, $df = 1$, $p = 0.005$; shoes: PT interaction model against nullmodel $D = 10.15$, $df = 2$, $p = 0.006$). This difference was to be expected and can be attributed to the experimental setup: For CP participants, the presentation times were individualized based on the participants performance in recognizing frontal images; for control participants the assignment of presentation times was independent of individual performance. The individualized presentation times for frontal face images appear adequate to capture individual variability among CP participant also for rotated images: While control participants with longer presentation times performed better (faces: $\hat{\beta} = 0.45$, $\text{HPDI}_{95\%} = [0.20, 0.57]$; shoes: $\hat{\beta} = 0.71$, $\text{HPDI}_{95\%} = [0.53, 0.90]$), there was no effect of presentation time on performance for CP participants in recognizing faces ($\hat{\beta} = 0.02$, $\text{HPDI}_{95\%} = [-0.22, 0.25]$) and only a slight effect in recognizing shoes ($\hat{\beta} = 0.24$, $\text{HPDI}_{95\%} = [0.05, 0.50]$). The influence of rotation on recognition performance didn't differ between CP and control participants neither for faces (LR-test of full interaction model against PT interaction model; $D = 4.64$, $df = 3$, $p = 0.20$) nor shoes ($D = 3.60$, $df = 2$, $p = 0.17$).

Comparing residuals, CP participants performed worse in the recognition of rotated images for faces (Wilcoxon rank sum, $W = 400$, $n_0 = 30$, $n_{CP} = 15$, $p < .001$ one-sided) but not for shoes ($W = 217$, $n_0 = 32$, $n_{CP} = 16$, $p = 0.58$ one-sided). The difference in performance between face and shoe recognition was selective for all CP participants (see figure 1 F).

Experiment 5

In experiment 5 we tested face recognition for stimuli with a limited presentation time during initial encoding. Compared to controls, CP participants made more mistakes in recognizing target faces. Interestingly, this difference in performance was already present for presentations of only 50 ms and didn't change across different presentation times. In contrast to previous results that controls responded faster than CP participants if presentation time is unlimited, under limited presentation there was no difference in reaction times. However, while controls responded faster if the target has been previously presented longer, no such relation could be observed among CP participants.

On average control participants performed better than CP participants for all of the presentation times tested (Wilcoxon rank sum test, all $p < 0.05$, see figure 2 B). Model based comparisons revealed a significant main effect of group (LR test of of main effect: $D = 13.3$, $df = 1$, $p < .001$). Although both groups improved in performance with increasing presentation time during the learning of the stimuli, the performance of control participants increased slightly stronger than the performance of CP participants (LR-test of full model against main effect model, $D = 4.06$, $df = 1$, $p = 0.044$), with a lower increase in performance for increasing presentation times in CP participants compared to controls ($\hat{\beta} = -0.13$, $\text{HPDI}_{95\%} = [-0.23, 0.01]$). Thus, control participants profited more from increased presentation times during learning than CP participants.

There was no difference in average reaction time between control and CP participants for any of the presentation times (Wilcoxon rank sum test, all $p > 0.1$ two-sided, see figure 2 C). This was confirmed by a model comparison (LR-test of main effect: $D = 0.32$, $df = 1$, $p = 0.57$). However, a significant difference in the influence of presentation time during learning on reaction time during recognition was found (LR-test of full model against null model: $D = 9.94$, $df = 2$, $p = 0.006$): While controls participants' reaction times decreased with increasing learning time ($\hat{\beta} = 3.6 \cdot 10^{-5}$, $\text{HPDI}_{95\%} = [2.5, 4.7] \cdot 10^{-5}$, note that coefficients are with respect to an inverse scale), there was no significant influence for CP participants ($\hat{\beta} = 1.13 \cdot 10^{-5}$, $\text{HPDI}_{95\%} = [-0.8, 2.2] \cdot 10^{-5}$, inverse scale).

Residuals with respect to the nullmodel are greater for CP participants compared to control participants in recognition performance (Wilcoxon rank sum test, $W = 261$, $n_0 = 24$, $n_{CP} = 13$, $p < 0.001$ one-sided) but there's no difference in residuals of reaction times ($W = 161$, $n_0 = 24$, $n_{CP} = 13$, $p = 0.89$ two-sided).

Experiment 6

Experiment 6 assessed participants performance in recognizing faces that were presented for a limited duration during the recognition phase. Similar to experiment 5, CP participants made more mistakes in recognizing target faces than controls. Again, this difference in performance was already present for presentations of only 50 ms but compared to experiment 5 the difference didn't increase with increasing different presentation time. There were no differences between CP participants' and controls' reaction times.

On average control participants performed better than CP participants already after a presentation of only 50 ms (Wilcoxon rank sum test, all $p < 0.05$, see figure 2 E). Model based

comparisons revealed a significant main effect of group (LR test of of main effect: $D = 10.03$, $df = 1$, $p = .002$). Both groups improved in performance with increasing presentation time ($\hat{\beta} = 0.17$, $\text{HPDI}_{95\%} = [0.14, 0.21]$), without a significant difference (LR-test of full model against main effect model, $D = 0.48$, $df = 1$, $p = 0.49$).

There was no difference in average reaction time between control and CP participants for any of the presentation times (Wilcoxon rank sum test, all $p > 0.3$ two-sided, see figure 2 F). This was confirmed by a model comparison (LR-test of main effect: $D = 0.03$, $df = 1$, $p = 0.87$). There was no significant difference in the influence of presentation time on reaction time between the groups (LR-test of full model against null model: $D = 0.87$, $df = 2$, $p = 0.65$).

Residuals with respect to the nullmodel are greater for CP participants compared to control participants in recognition performance (Wilcoxon rank sum test, $W = 255$, $n_0 = 24$, $n_{CP} = 13$, $p < 0.001$ one-sided) but there's no difference in residuals of reaction times ($W = 174$, $n_0 = 24$, $n_{CP} = 13$, $p = 0.58$ two-sided).

Discussion

In all of the face recognition tests there was a significant difference in performance between the group of CP participants and the control group: CP participants needed more initial training, had longer reaction times and needed longer presentation times of stimuli to achieve the same level of accuracy as compared to controls. The face recognition deficit in CP participants was present for both: frontal views of the faces, which were extensively trained, and rotated views of the faces, which were only presented as test stimuli. In contrast to the recognition of faces, there was no performance difference between the two groups in discriminating individual shoes. However, for four CP participants (HE,SE,HB,RK) the deficits seemed to extent to object identification.

Regarding out hypothesis, we first replicated findings that CP participants can achieve a face recognition accuracy comparable to controls, albeit requiring more initial training and longer reaction times. This increase in reaction time was paralleled by an equal-sized increase in the presentation time CP participants needed in order to perform at the same level as controls. Furthermore, if stimulus presentation is limited during recognition (experiment 6), these differences in reaction time during recognition vanish. Thus, CP participants can achieve normal recognition accuracy but they need to inspect the stimulus longer.

Second, CP participants showed worse recognition accuracy than controls if presentation time - and therefore the process of information extraction - was limited (experiments 5 and 6). The difference was already present for presentation times of 50ms. Thus, performance differences are not merely a function of slower processing but of differences in the processing that are present from the very beginning.

Third, for both groups, the influence of limited presentation time on later recognition accuracy was more pronounced if limitation took place during learning compared to recognition of faces. However, group differences in this positive influence were only present for presentation times limited during the initial encoding (experiment 5). Furthermore, the positive influence of increased encoding time extended to a decrease in reaction times among controls but not among CP participants. Thus, for controls recognition accuracy and reaction time were limited

to a larger degree than for CP participants by restricted inspection time during initial encoding than during the recognition itself. Whereas controls are able to refine their encoding of a face by incorporating more and more information (possibly up to the point of very fast and highly accurate recognition) there seem to be more stringent limits to the information capacity of facial encoding in congenital prosopagnosia.

Fourth, performance in face recognition decreased with increasing rotation angle between learned view and testing view for both CPs and controls similarly. Although this finding supports the original model put forward, the absence of group differences in the influence of rotation might have been due to different reasons. First, in the experiments participants were trained only on frontal images. However, it is questionable whether frontal images are suited to the construction of 3D models [36] and whether the construction can be based on the observation of a single image at all. Second, isolated features might posses a certain inherent degree of transformation invariance against experimental stimulus manipulations. For example, one CP participant used cues with a high degree of rotation invariance as part of their compensatory processing or feature selection strategies. In recognizing one of the target faces, FP attended a small mole placed on the left side of the neck which was visible in frontal view and for rotations to the left but hidden for rotations to the right. Accordingly FP recognized this target face whenever it was rotated to the left and failed to recognize the target whenever it was rotated to the right, irrespective of the rotation angle.

The cognitive heterogeneity in congenital prosopagnosia (cf. [23]) raises the question whether there are identifiable subgroups of CP, comparable to those found in acquired prosopagnosia [6]. In acquired prosopagnosia (AP) a mature, fully functional face recognition system is disturbed by an external event, unrelated to the system's past performance. Irrespective of the exact processes underlying functional specialization of cortical regions in the neural system of face recognition, this specialization presumably leads to an alignment between cortical location and functional process [56–59]. Damage inflicted to a specific region can therefore lead to restricted deficits, conditional on the interconnectedness and interdependence of the distributed processing [60]. However, in contrast to the acquired form, individuals with congenital prosopagnosia never evolve a functional face recognition system in the first place and their deficit has to be interpreted as an endpoint of a developmental trajectory [61], a mature but dysfunctional system. Thus, even if there is a single initial cause to CP, it would not be surprising to see a stronger heterogeneity in CP compared to a homogeneous group of AP participants, i.e. with the same lesions, based solely on differences in development, e.g. learning of different evasive and compensatory strategies. Since the strategies adopted vary greatly between individual CPs [16, 19], this complicates a categorization of the intrinsic heterogeneity in CP based on a small number of behavioral tests. Thus, in future studies it seems essential to integrate behavioural as well as neurophysiological/-anatomical variability with computational models of CP based on a general theory of visual information processing.

Acknowledgments

The authors thank all control and PA subjects for their participation, Isabelle Bülthoff (MPI Biol. Cyb.), Tobias Wolf (MPI Biol. Cyb.), Brigitte Welling (Uni. Münster) Bernhard Englitz

(MPI MiS) for interesting discussions and technical support.

References

References

1. Bodamer J (1947) Die Prosop-Agnosie. Arch Psychiatr Nervenkr 179: 6–53.
2. Quaglino A, Borelli GB, Della Sala S, Young AW (2003) Quaglino's 1867 case of prosopagnosia. Cortex 39: 533–40.
3. Lissauer H (1890) Ein Fall von Seelenblindheit nebst einem Beitrag zur Theorie derselben. Arch Psychiatr Nervenkr 21: 22–70.
4. Wilbrand H (1892) Ein Fall von Seelenblindheit und Hemianopsie mit Sectionsbefund. Dtsch Z Nervenheilk 2: 361–387.
5. Hecaen H, Angelergues R (1962) Agnosia for faces (prosopagnosia). Arch Neurol 7: 92–100.
6. Damasio A, Tranel D, Damasio H (1990) Face Agnosia and the Neural Substrates of Memory. Annu Rev Neurosci 13: 89–109.
7. De Renzi E, Faglioni P, Grossi D, Nichelli P (1991) Apperceptive and associative forms of prosopagnosia. Cortex 27: 213–21.
8. De Renzi E, Perani D, Carlesimo G, Silveri M, Fazio F (1994) Prosopagnosia can be Associated with Damage Confined to the Right-Hemisphere - An MRI and PET Study and a Review of the Literature. Neuropsychologia 32: 893–902.
9. Barton JJS, Cherkasova M, O'Connor M (2001) Covert recognition in acquired and developmental prosopagnosia. Neurology 57: 1161–1168.
10. McConachie HR (1976) Developmental prosopagnosia. A single case report. Cortex 12: 76–82.
11. Ariel R, Sadeh M (1996) Congenital visual agnosia and prosopagnosia in a child: a case report. Cortex 32: 221–40.
12. Kress T, Daum I (2003) Developmental prosopagnosia: a review. Behavioural neurology 14: 109–21.
13. Hasson U, Avidan G, Deouell LY, Bentin S, Malach R (2003) Face-selective activation in a congenital prosopagnosic subject. J Cognitive Neurosci 15: 419–31.
14. Behrmann M, Avidan G (2005) Congenital prosopagnosia: face-blind from birth. Trends in Cognitive Sciences 9: 180–7.

15. Behrmann M, Avidan G, Marotta J, Kimchi R (2005) Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. J Cognitive Neurosci 17: 1130–1149.

16. Kennerknecht I, Grüter T, Welling B, Wentzek S, Horst J, et al. (2006) First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). Am J Med Genet 140A: 1617–1622.

17. Grüter M, Grüter T, Bell V, Horst J, Laskowski W, et al. (2007) Hereditary Prosopagnosia: the First Case Series. Cortex 43: 734–749.

18. Kennerknecht I, Plümpe N, Edwards S, Raman R (2007) Hereditary prosopagnosia (HPA): the first report outside the Caucasian population. J Hum Genet 52: 230–6.

19. Kennerknecht I, Pluempe N, Welling B (2008) Congenital prosopagnosia - a common hereditary cognitive dysfunction in humans. Front Biosci 13: 3150–3158.

20. Duchaine BC (2006) Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. Cognitive Neuropsychology 23: 714–747.

21. Lange J, De Lussanet M, Kuhlmann S, Zimmermann A, Lappe M, et al. (2009) Impairments of Biological Motion Perception in Congenital Prosopagnosia. PLoS ONE 4: e7414.

22. Grueter T, Grueter M, Bell V, Carbon CC (2009) Visual mental imagery in congenital prosopagnosia. Neurosci Lett 453: 135–140.

23. Schmalzl L, Palermo R, Coltheart M (2008) Cognitive heterogeneity in genetically based prosopagnosia: A family study. J Neuropsychol 2: 99–117.

24. Barton JJS, Zhao J, Keenan JP (2003) Perception of global facial geometry in the inversion effect and prosopagnosia. Neuropsychologia 41: 1703–11.

25. Farah MJ, Wilson KD, Drain M, Tanaka JN (1998) What is "special" about face perception? Psychol Rev 105: 482–98.

26. Blais C, Jack RE, Scheepers C, Fiset D, Caldara R (2008) Culture shapes how we look at faces. PLoS ONE 3: e3022.

27. Bindemann M, Burton AM, Scheepers C (2008) How are frontal, mid-profile, and profile faces processed? Evidence from eye movements. Perception 37: 33.

28. Henderson JM, Williams CC, Falk RJ (2005) Eye movements are functional during face learning. Memory & cognition 33: 98–106.

29. Schwarzer G, Huber S, Grüter M, Grüter T, Gross C, et al. (2007) Gaze behaviour in hereditary prosopagnosia. Psychological Research 71: 583–90.

30. Schmalzl L, Palermo R, Green M, Brunsdon R, Coltheart M (2008) Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. Cognitive Neuropsychology 25: 704–729.

31. Duchaine BC, Nakayama K (2005) Dissociations of Face and Object Recognition in Developmental Prosopagnosia. J Cognitive Neurosci 17: 249–261.

32. Faust C (1947) Partielle Seelenblindheit nach Occipitalhirnverletzung mit besonderer Beeinträchtigung des Physiognomieerkennens. Der Nervenarzt : 294–297.

33. Gauthier I, Behrmann M, Tarr MJ (1999) Can face recognition really be dissociated from object recognition? J Cognitive Neurosci 11: 349–370.

34. Shepherd J, Deregowski J, Ellis H (1974) A Cross-Cultural Study of Recognition Memory For Faces. International Journal of Psychology 9: 205 – 212.

35. Avidan G, Hasson U, Malach R, Behrmann M (2005) Detailed exploration of face-related processing in congenital prosopagnosia: 2. Functional neuroimaging findings. J Cognitive Neurosci 17: 1150–1167.

36. Troje NF, Bülthoff HH (1996) Face recognition under varying poses: the role of texture and shape. Vision Res 36: 1761–71.

37. Duchaine BC, Nakayama K (2006) The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia 44: 576–85.

38. Grüter M (2004) Genetik der kongenitalen Prosopagnosie [Genetics of congenital prosopagnosia]. Thesis (MD), Münster University.

39. Kennerknecht I, Ho NY, Wong VCN (2008) Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. Am J Med Genet 146A: 2863–70.

40. Riddoch JM, Humphreys GW (1993) BORB: Birmingham Object Recognition Battery. Psychology Press, NY.

41. Warrington EK, James M (1991) The Visual Object and Space Perception Battery. Harcourt Assessment, London.

42. Kesten H (1958) Accelerated Stochastic-Approximation. Ann Math Stat 29: 41–59.

43. Treutwein B (1995) Adaptive Psychophysical Procedures. Vision Res 35: 2503–2522.

44. Farkas LG (1981) Anthropometry of the Head and Face in Medicine. Elsevier, New York, 293 pp.

45. Munoz D, Broughton J, Goldring J, Armstrong I (1998) Age-related performance of human subjects on saccadic eye movement tasks. Experimental Brain Research 121: 391–400.

46. Elze T (2009) FlashDot - A platform independent experiment generator for visual psychophysics. JOV 9: 58–58.

47. Elze T (2006) Shorter than you think? Temporal precision of visual stimuli on computer monitors: Pitfalls and systematic errors. Perception 35: 52–52.

48. R Development Core Team (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

49. Wichmann FA, Hill N (2001) The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys 63: 1293–1313.

50. Wichmann FA, Hill N (2001) The psychometric function: II. Bootstrap-based confidence intervals and sampling. Percept Psychophys 63: 1314–1329.

51. Faraway JJ (2006) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Chapman & Hall, FL : 301.

52. Bates D, Maechler M (2009) lme4: Linear mixed-effects models using S4 classes. URL `http://CRAN.R-project.org/package=lme4`. R package version 0.999375-31.

53. Hadfield J (2009) MCMC methods for Multi-response Generalised Linear Mixed Models: The MCMCglmm R Package. submitted 1: 1.

54. Baayen R, Davidson D, Bates D (2008) Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language .

55. Ratcliff R (1993) Methods for dealing with reaction time outliers. Psychological Bulletin 114: 510–32.

56. Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17: 4302–11.

57. Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, et al. (2000) The fusiform "face area" is part of a network that processes faces at the individual level. J Cognitive Neurosci 12: 495–504.

58. Hoffman E, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci 3: 80–84.

59. Haxby JV, Hoffman E, Gobbini M (2000) The distributed human neural system for face perception. Trends in Cognitive Sciences 4: 223–233.

60. Fox CJ, Iaria G, Barton JJS (2008) Disconnection in prosopagnosia and face processing. Cortex 44: 996–1009.

61. Thomas M, Karmiloff-Smith A (2003) Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. Behavioral and Brain Sciences 25: 727–750.

Tables

Initials	Contact	Age	Gender
HE	Self-reported	68	F
SE	Self-reported	64	F
EB	Self-reported	57	F
HG	Self-reported	53	M
HB	Self-reported	50	M
MB	Self-reported	48	F
MR	Self-reported	48	F
RK	Self-reported	41	M
JM	Self-reported	26	M
JF	Screening	23	M
HS	Screening	22	M
VK	Screening	21	M
FP	Self-reported	21	F
MG	Screening	21	F
HW	Self-reported	21	F
LL	Screening	21	F

Table 1. Description of CP participants. Age is with respect to November 2006 (first series of experiments)

Figures

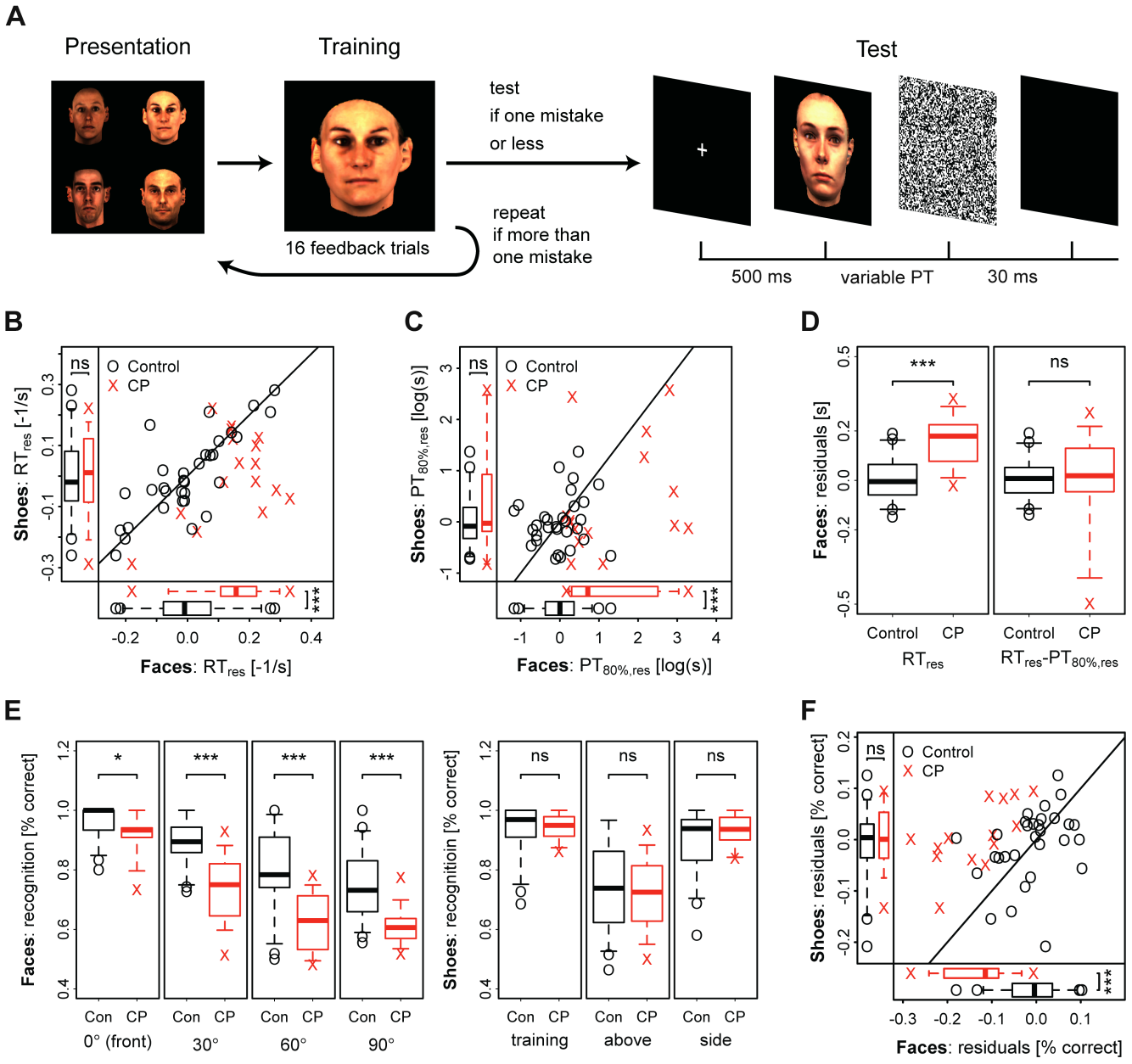

Figure 1. Contrasting face and shoe recognition for frontal and rotated images. (**A**) In experiments 1-4 participants were first presented with 4 target stimuli, and were trained during at least 16 feedback trials prior to the test. (**B**) During feedback training, participants with congenital prosopagnosia (CP) had longer reaction times (RT_{res}, inverse transformation i.e. $-1/RT$ see Methods) compared to controls for face stimuli (prior to experiments 1 and 3) but not for shoe stimuli (prior to experiments 2 and 4). (**C**) CP participants needed longer presentation times to achieve 80% correct recognition ($PT_{80\%,res}$) than controls in tests for face recognition (Exp. 2) but not for shoe recognition (Exp. 3). (**D**) Group differences in RT_{res} for faces stimuli vanished after accounting for differences in $PT_{80\%,res}$: CP participants needed to inspect face stimuli longer than controls. In the recognition of rotated images, CP participants performed worse than controls for for faces (Exp. 4) but not for shoes (Exp. 5); both for every rotation angle separately (**E**) as well as in individual residuals which account for differences in the experimental setup (**F**). Note that the group difference observed for face recognition didn't change with rotation angle (Likelihood-ratio test, $p = 0.2$). In all figures, scatter-plots depict average residuals for each individual obtained after fitting GLMMs, and boxplots show group distributions (whiskers: 90% CI) for each stimulus/condition separately complemented by significance values according to a Wilcoxon rank sum test.

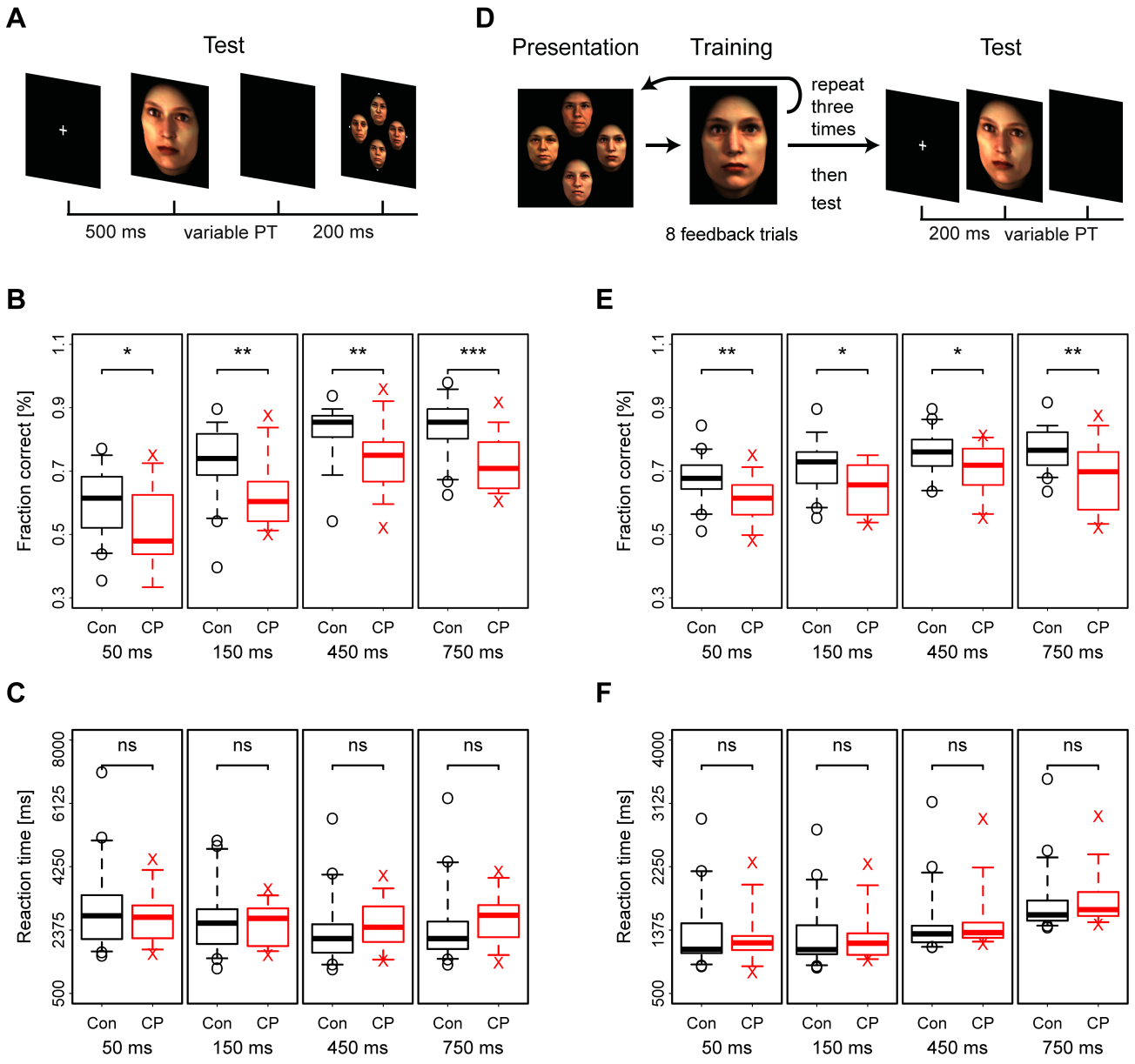

Figure 2. Face recognition under the constraint of limited presentation time. (**A**) In experiment 5, participants were shortly presented with a target stimulus for 50ms, 150ms, 450ms or 750ms, and after a short blank had to recognize the target in a display of four face images. For all presentation times used in experiment 5, performance of CP participants differed from controls (**B**), while there was no significant difference in reaction times (**C**). However, while controls responded faster with increasing presentation time, there was no significant influence of presentation time on CP reaction times (Likelihood-ratio test, $p = 0.006$). (**D**) In experiment 6, participants were repeatedly presented target faces, which - after a total of 24 feedback trials - had to be recognized in a 2-alternative forced choice paradigm. During the test, faces were presented for variable durations (50ms, 150ms, 450ms or 750ms). Independent of the duration, CP participants performed worse than controls (**E**), without significant differences in reaction times (**F**).