LETTERS
https://doi.org/10.1038/s41562-017-0266-3

nature human behaviour

Corrected: Publisher correction

The determinants of consciousness of human faces

Yaniv Abir[1*], Asael Y. Sklar[2], Ron Dotsch[3], Alexander Todorov[4] and Ran R. Hassin[5*]

From what we see to what we hear and from how we feel to what we think, our conscious experiences play an important role in shaping our lives. Because we become aware of only a small subset of our ongoing cognitive and perceptual processes[1–4], explicating the determinants of conscious experiences is a crucial step towards understanding human behaviour. Here we develop a computational data-driven approach for studying the determinants of consciousness and we use it to investigate what is arguably the most important social stimulus: the human face[5–7]. In six experiments with 174 participants, we used this method to uncover a reliable dimension that determines the speed with which different faces reach conscious awareness. This dimension correlates strongly with the perceived power/dominance of a face. We show that the dimension cannot be explained by low-level visual factors and does not describe conscious processing, thereby suggesting that it captures the process of prioritization for consciousness. By visualizing the dimension, we are able to produce a vivid depiction of what unconscious processes prioritize for conscious processing. We propose this method as a means to study the contents and neural correlates of conscious experiences across various domains.

Research into the factors determining conscious awareness is often conducted by examining how fast we become aware of various visual stimuli under different conditions (but see other paradigms[4,8–11]). Presentation techniques in which awareness can be dissociated from other cognitive processing have been fruitfully used to this end. Chief among these is the recently developed continuous flash suppression (CFS)[12]. In CFS a target stimulus is rendered invisible by presenting it to one eye, while a high-contrast flickering suppressor is presented to the other eye. Presentation parameters can be tuned for the target stimulus to become visible after several hundreds of milliseconds (or even longer). Participants in such experiments are asked to respond as soon as they become aware of the stimulus (for example, by indicating whether it is above or below fixation—a paradigm called breaking CFS or bCFS; Fig. 1a)[13]. Response times—or breaking times (BTs)—are a measure of how long it takes participants to become aware of the stimulus, and are therefore also a measure of prioritization for consciousness.

Given the social importance of faces, it is of little surprise that they have been widely used to examine prioritization for consciousness[5]. Faces enjoy privileged processing in the cognitive system[14], be it conscious or non-conscious, and their processing is supported by specific neural circuitry[15–17]. It is thus not surprising that faces break into consciousness faster than control stimuli composed of the same low-level visual features[13,18,19]. Factors distinguishing different faces have also been found to modulate BTs, among them emotional expression[20], gaze direction[21,22], personal familiarity[23] and perceived social traits[24].

Unlike any previous study of prioritization for consciousness, of faces or otherwise, we have developed a computational, data-driven approach for the delineation of the determinants of conscious experience. In a typical bCFS experiment, stimuli are selected according to the researchers' hypothesis, and BTs for these stimuli are compared across a dimension(s) of interest. While allowing tight experimental control, this approach may constrain possible empirical outcomes and their ecological validity in various ways. First, hypothesis-driven experiments that address similar questions are limited so that only pre-hypothesized dimensions can ever be examined[6,25]. Additionally, the cognitive system itself might be biased by properties of the chosen stimulus set[26,27]. The use of a data-driven approach with randomly generated stimuli (see, for example, refs [6,7,28–31]) allows us to overcome these limitations and construct a rich model of the relevant dimensions[6,28]. Moreover, the data-driven approach can discover models that constitute new hypotheses, which can then be rigorously tested in a hypothesis-driven approach. Hence, the two methods should be viewed as complementary.

In the experiments we report here, we used a set of 300 computer-generated face images. The faces were derived from a statistical face-space model that captures the variance from a large sample of real faces. This face-space has 50 orthogonal dimensions that define the shape and 50 orthogonal dimensions that define the reflectance of a face stimulus. Our 300 stimuli were randomly drawn from a normal distribution on each of these parameters[7,25].

In experiment 1, BTs for each of the 300 faces were measured in a sample of 25 participants. We used a mirror stereoscope to separate the visual input to the two eyes. On each trial, one of the 300 faces was presented to one eye, while a multicolour, high-contrast dynamic pattern of rectangles was presented to the other eye. Participants were instructed to indicate with a key press, as quickly as possible, whether the face appeared above or below fixation (Fig. 1a). Given that this is a simple perceptual task, this measure allowed us to estimate how long it takes for a participant to become conscious of a face. On average, participants became aware of the faces after 1,439 ms (s.d. = 368 ms).

We then used reverse correlation to model the facial properties predictive of BT. The vector of parameters defining the shape of each of the stimulus faces was multiplied by the BT measured for that face. These vectors were then averaged, producing a dimension[25] predictive of how long it will take for a face to become conscious. The resulting dimension was scaled such that the within-participant

[1]Department of Cognitive Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [2]Department of Psychology, The Ohio State University, Columbus, OH, USA. [3]Department of Psychology, Utrecht University, Utrecht, Netherlands. [4]Department of Psychology, Princeton University, Princeton, NJ, USA. [5]Department of Psychology and The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem, Israel. *e-mail: yaniv.abir@mail.huji.ac.il; ran.hassin@huji.ac.il

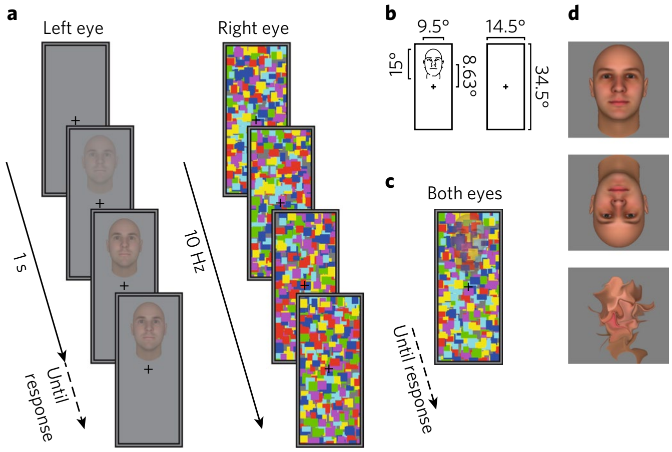

Fig. 1 | Experimental paradigm and examples of stimuli. a, Schematic of a breaking continuous flash suppression trial. The left frame is presented to the left eye, the right frame to the right eye. **b**, Visual angle of stimuli presentation. **c**, Schematic of a conscious RT control trial: face image blended into a Mondrian pattern. For visualization purposes, face image is reproduced here at 60% contrast; in the experiment, faces were blended at 35% contrast. **d**, Stimuli examples. Top to bottom: upright face, experiments 1, 2 and 6; inverted face, experiment 5; diffeomorphically scrambled face, experiment 4.

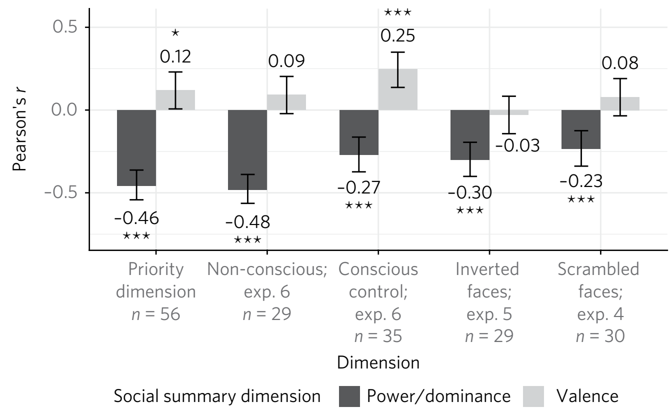

Fig. 3 | Correlations with the two summary dimensions of judgements of social traits. Coefficients for the correlations of predicted BTs from different experiments with the ratings for the two summary dimensions of judgements of social traits. Error bars denote correlation 95% confidence interval. $^*P<0.05$; $^{***}P<0.001$.

z scores of the measured BTs were used as its units. This dimension correlated with BTs in this first sample, $r=0.44$, $P<0.001$, explaining 19.36% of the variance in BTs. Thus, in experiment 1 we identified a dimension in face-space that correlates with prioritization for conscious awareness. We examined the causal nature of these relationships in experiment 3.

Experiment 2 was a replication with another sample of 31 participants. Crucially, the model derived from experiment 1 significantly predicted BTs in experiment 2, and vice versa (average $r=0.31$, $P<0.001$). Collapsing over both samples, relative BTs were consistent over the 56 participants (interclass correlation: ICC(2,56) = 0.67)[32] and the extracted dimension explains 24.15% of the variance in BTs. We will henceforth refer to this dimension as the priority dimension.

A quick glance at the visualization of the priority dimension (Fig. 2 and Supplementary Video) suggests that it is socially meaningful. To analytically examine the social meaningfulness of the priority dimension, we correlated it with the two central social traits inferred from faces: valence/trustworthiness and power/dominance[7]. The priority dimension is significantly correlated with power/dominance scores, so faces judged as more powerful break earlier into consciousness, $r=-0.46$, $P<0.001$. The priority dimension showed a weaker correlation with valence/trustworthiness scores, such that untrustworthy faces break earlier into consciousness $r=0.12$, $P=0.04$ (Fig. 3).

Experiments 1 and 2 revealed a correlation between the priority dimension and the timing of conscious experience. To experimentally manipulate and validate this dimension, in experiment 3 ($n=30$) we measured BTs for faces varying only in the priority dimension. The faces ranged from $z=-5$ to $z=5$ on the dimension, in 21 equal steps. A planned linear contrast on the results showed that the value of a face on the priority dimension was a significant determinant of BTs, $F(1,29)=7.27$, $P=0.012$, partial $\eta^2=0.20$ (Fig. 4; see Methods for a convergent mixed-models analysis).

To determine whether the priority dimension can be explained by low-level visual factors known to affect BTs (such as contrast or spectral content), we conducted two control experiments ($n=30$ in experiment 4, $n=29$ in experiment 5). In experiment 4 we measured relative BTs for diffeomorphically scrambled versions of the 300 faces—a method that preserves low-level qualities while rendering the stimulus unidentifiable (Fig. 1b)[33]. In experiment 5 we measured BTs for the 300 faces presented upside down, thus preserving all low-level visual features, but impeding holistic face-processing (Fig. 1d)[14]. We then repeated the dimension extraction procedure with BTs for the scrambled and inverted faces. To estimate the proportion of variance in the priority dimension model

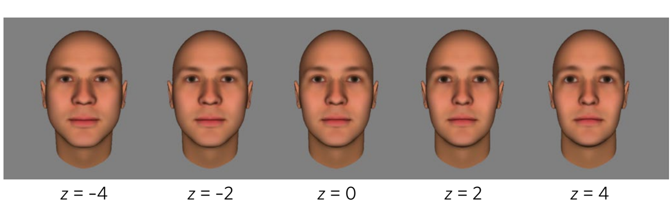

Fig. 2 | Visualization of the priority dimension. Predicted BTs increase from left to right.

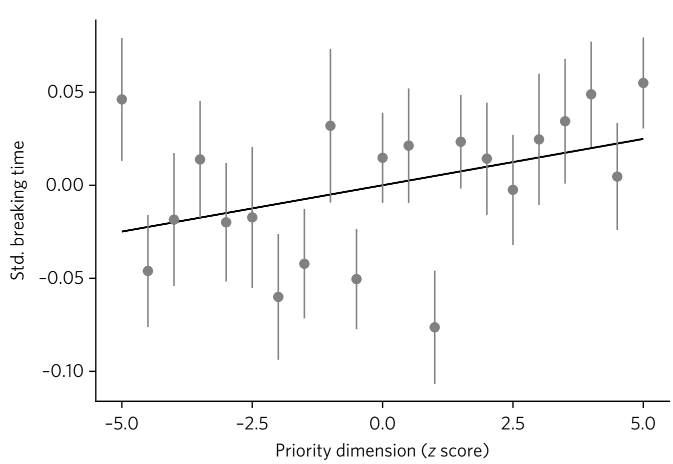

Fig. 4 | Experiment 3 results. Average standardized BTs for 21 levels on the priority dimension ($n=30$). BTs were standardized per participant before averaging. Error bars denote s.e. across participants. The linear regression line is also plotted.

explained by low-level features, we conducted a partial-correlation analysis. Crucially, the dimension extracted from experiment 2 significantly predicted BTs in experiment 1, even when controlling for the low-level features dimension extracted from the scrambled faces (partial $r=0.29$, $P<0.001$), non-holistic processing dimension extracted from the inverted faces (partial $r=0.17$, $P=0.004$) and when controlling for both (partial $r=0.15$, $P=0.01$). Similar results were found when computing partial-correlations from the dimension extracted in experiment 1 to BTs from experiment 2 (see Methods). Thus, low-level features are not sufficient to explain the variance predicted by the priority dimension. Interestingly, the low-level dimensions did correlate with power/dominance scores (Fig. 3), indicating that some of the relation with power/dominance may be driven by visual features preserved in the control stimuli.

Finally, in experiment 6 we examined whether the priority dimension describes a more general variation in conscious response times (RTs) to faces[13]. As in the original bCFS task (experiments 1 and, 2), 35 participants were asked to indicate as soon as they knew whether each of the 300 faces was above or below fixation. In this task, however, faces were presented to both eyes, blended with a static multicolour suppressor image (Fig. 1c). Thus, the faces could be consciously perceived from the outset of each trial. At the end of this task, participants completed the original bCFS task.

We extracted the dimension that underlies performance in the conscious part of this experiment, as well as the dimension that underlies non-conscious processes. If the priority dimension uncovered in experiments 1 and 2 reflects conscious processes, then the conscious dimension should completely mediate the relationship between the priority dimension and BTs. However, the priority dimension significantly predicted BTs in this experiment, even when controlling for the dimension extracted from conscious RTs, $r=0.26$, $P<0.001$. Importantly, the priority dimension was a significantly better predictor of BTs in this experiment ($r=0.29$, $P<0.001$) than of conscious RTs ($r=0.12$, $P=0.04$; Williams's $t(297)=2.21$, $P=0.03$).

The nature of the correlations of the two dimensions with dimensions derived from explicit personality judgements provides evidence for their dissociability and hence for the non-conscious nature of the priority dimension. Specifically, the conscious dimension had a significantly weaker correlation with power/dominance scores (conscious RTs: $r=-0.27$, $P<0.001$; BTs: $r=-0.48$, $P<0.001$; Williams's $t(297)=3.71$, $P<0.001$) and a stronger correlation with the valence scores ($r=0.25$, $P<0.001$; Williams's $t(297)=2.48$, $P=0.014$).

In six experiments we have developed and tested a computational data-driven method for modelling prioritization for consciousness. Using this method with our specific database of faces, we have discovered a reliable dimension that determines the speed with which these faces become conscious. The dimension is socially meaningful in that it correlates with the perceived power/dominance of a face. These results imply that prioritization for consciousness of faces is a functionally evaluative process, operating on socially relevant information.

Methodologically, we have presented here a potent method to describe and visualize the operation of prioritization for consciousness. We have reported its use with one stimulus category—human faces—but this method can be applied broadly to other domains of interest, to uncover dimensions that prioritize conscious experiences. Any stimulus class that can be formalized as a multidimensional space of attributes, such as body posture[34], facial expression[6] or semantic meaning[35], would be suitable for such undertaking.

Our study describes the components of the priority dimension that are common among our participants. A promising avenue for research is to examine individual differences in the priority dimension and their relation to other psychological and physiological traits[36,37]. Although our experiments were not designed for this purpose, the promising test–retest reliability range in our data ($r=0.29$–0.54, $P<0.001$; see Methods) is supportive of such a possibility. Such research would also allow for an examination of potential gender differences in the priority dimension, a question left open because of the predominance of women in our participant population. Similarly, the dimension may (and we believe is likely to) be affected by contexts, both external (for example, the distribution of faces in the world) and internal (for example, motivation). Thus, the adaptivity of the dimension to changing circumstances is another promising avenue for future research.

This Letter leaves, of course, many other open questions. One of them is whether people become aware of the whole face or only parts of the face. Recall that we measured the time any one part of the face took to become visible, and we therefore cannot be certain about the content of our participants' consciousness. The nature of the priority dimension for whole faces (one that allows, for example, for processes such as identification) is an exciting question left for future investigations. Similarly open is the question regarding the role of facial dimensions involved in simple conscious RT tasks, like the one in experiment 6.

Finally, our paper suggests that the reverse correlation method could be used as a sensitive tool for the study of conscious experiences. In this Letter we have demonstrated its ability to study, model and visualize face prioritization for conscious processing and its comparison with conscious response times and dimensions. In the future, it could serve as a common denominator for the comparison and integration of a wide range of once disparate methods for measuring conscious experience processes and their neural correlates. Once reverse correlation models of brain activation patterns and of attentional and perceptual responses to faces are reliably built, they can be directly correlated and compared. The modelled dimensions could thus act as bridges between measures of consciousness emanating from different physical and psychological levels of explanation.

Methods

The protocol for these experiments was approved by the Institutional Review Board of the Psychology Department of The Hebrew University of Jerusalem. Informed consent was obtained from all participants.

Experiment 1. *Participants.* We pre-specified a sample size of at least 30 participants for this experiment on the basis of previous studies applying reverse correlation to ratings of faces[7]. Hebrew University students ($n=31$, 19 women; age: mean, $M=23.75$, s.d. $=3.48$) participated in return for course credit ($n=2$) or payment of 30 NIS. Of the participants, 30 reported normal vision and one had corrected to normal vision with contact lenses.

Materials. Stimuli consisted of multicolour suppressors ('Mondrians'—random amalgams of partly overlapping rectangles of varying sizes and colours) and target stimuli. The target pictures were 300 faces generated using FaceGen 3.1 and adapted from previous studies[7]. A total of 62 additional face stimuli were generated for the preliminary stage, training block and a familiarity rating task.

Stimuli were presented on an Eizo FlexScan F520 CRT monitor (refresh rate 100 Hz), using PsychToolbox[38] for MATLAB. Participants viewed the stimuli through a mirror stereoscope so that the visual input to each eye was controlled separately. A frame was presented around the stimuli to aid convergence of the visual percept. Viewing distance was ~32 cm. Figure 1 depicts the stimuli and the visual angle of presentation.

Procedure. Participants first read the instructions for the experiment. The experimenter adjusted the stereoscope to achieve convergence of the image presented to each eye of the participant. Participants performed a non-suppressed preliminary adaptation phase in which 25 faces were serially presented binocularly for 4 s each. Stimuli faded into view over the first second of presentation (maximum contrast of 35%). Eight randomly selected stimuli appeared twice consecutively. Participants had to indicate by pressing one of two keys whether the face presented was different from or identical to the previous one. The colour of the frame around the stimulus was then changed after the key press to provide feedback (green for correct, red for incorrect). Instructions for the adaptation task stated that the faces presented would be relevant at a later stage of the experiment.

After completing the preliminary phase, instructions for the bCFS task were presented on the screen. On each trial of the bCFS task, participants were presented with Mondrians changing at 10 Hz to one eye and one face stimulus to the other eye. The face image faded in during the first second of presentation (maximum contrast of 35%). Participants had to indicate by pressing one of two keys if the face appeared above or below fixation. They were instructed to respond as soon as they knew the location of the face. This reaction time, relative to trial onset, served as our dependent measure of BT. With these instructions, BTs measured prioritization for consciousness to any part of the face, allowing conservative inference regarding the non-conscious nature of the processes involved. If no response was given, the trial ended after 10 s. Stimuli order, presentation location and suppressed eye were randomized. Participants performed a training block of 25 stimuli, after which the experimenter left the room for the remainder of the task. Participants performed two blocks of the bCFS task, each with the 300 randomly generated faces (600 trials in total).

After completing the task, the stereoscope was removed. Twelve faces from the preliminary phase, 12 from the breaking suppression task and 12 new faces were consecutively presented in random order on screen. A scale of 1–7 appeared below each face and participants had to indicate using the keyboard how familiar the face was. This familiarity task was included so as to match the instructions given to participants during the adaptation phase and will not be discussed further. Participants also filled out a questionnaire, listing any special strategies employed during the experiment.

Statistics. Analyses for all experiments were carried out using the R statistical environment 3.3.2, with the reshape, plyr, ez, psych, ggm, bootstrap, pwr and lme4 packages. Outlier removal procedure were based on previous bCFS studies[39]. Three participants reported that they had shut one eye while performing the tasks. Their data were excluded from analysis. Mean accuracy of the experimental blocks was calculated per participant. As is the standard procedure when analysing bCFS, data from participants with accuracy below 90% were also excluded from analysis ($n = 3$), leaving the data of 25 participants in the final sample. Data from incorrect trials were excluded from analysis ($n = 402$, 2.68%). No BT below 200 ms was recorded in the remaining data. Data from trials in which the BT was more than 3 s.d. from the participant's mean were excluded from analysis ($n = 258$, 1.72%). BTs were then standardized per participant.

To estimate the reliability of the relative BTs, BTs from all participants were averaged for each presentation of each face. The correlation between the two presentations (test–retest reliability) was significant, $r = 0.40$, $P < 0.001$. Inter-rater agreement between participants was also estimated—BTs were averaged over presentations for each face and each participant, and an interclass correlation coefficient was estimated using a two-way random model, determining significant agreement for average measures ICC(2,25) = 0.43.

Following the methods in previous studies[7,25], we computed the reverse correlation dimension as the average of the 50 FaceGen shape parameters for each of the 300 faces, weighted by their corresponding relative BT across participants. Predicted BTs were computed from the dimension as the projection of each face on the dimension.

Experiment 2. *Participants.* Based on the participant inclusion criteria adopted in experiment 1, in this and all following experiments, we pre-specified a sample size of 30 participants after exclusion of participants by their questionnaire responses and overall performance. See Supplementary Methods for the power analyses. Exclusion rates were calculated at the end of each experimentation day, sometimes resulting in small variations in the number of participants. A total of 36 Hebrew University students participated in this experiment (28 women, age: $M = 22.38$, s.d. = 1.68) in return for course credit ($n = 29$) or payment ($n = 7$). Among these, 32 reported normal vision, and four had undergone laser eye surgery.

Materials and procedure. The stimuli, presentation and procedure were identical to those in experiment 1, except for exclusion of the familiarity rating task.

Statistics. Outlier removal criteria and data analysis methods were identical to those in experiment 1. Two participants reported shutting one eye during the experiment. One reported looking only at one side of the screen, pressing the key for the other side after a length of time. One participant reported losing convergence of the visual percept during the experiment. Their data were excluded from the analysis. Data from one participant with accuracy below 90% were also excluded from the analysis, leaving the data of 31 participants in the analysed sample. Data from incorrect trials were excluded from analysis ($n = 367$, 1.97%). No BT below 200 ms was recorded in the remaining data. Data from trials in which the BT was more than 3 s.d. from the participant's mean were excluded from analysis ($n = 315$, 1.69%).

The test–retest reliability for this sample was $r = 0.41$, $P < 0.001$. The inter-rater agreement was significant, ICC(2,31) = 0.56.

The reverse correlation dimension was extracted for this sample by the same method as used in experiment 1. The reverse correlation dimension was a significant predictor of BTs in this sample ($r = 0.48$, $P < 0.001$), explaining 23.04% of the variance in BTs. BTs from this experiment showed a significant correlation with BTs from experiment 1 ($r = 0.48$, $P < 0.001$). The correlations between predicted BTs from one experiment and actual BTs from the other were $r = 0.31$, $P < 0.001$ and $r = 0.32$, $P < 0.001$, indicating significant predictive ability between samples and convergence of the results from experiments 1 and 2. We therefore conducted analysis on the combined sample ($n = 56$). See Supplementary Fig. 1 for histograms of BTs in this participant sample.

The test–retest reliability of BTs over participants was $r = 0.54$, $P < 0.001$ in the combined sample. The final priority dimension used in the follow-up studies was computed based on this combined sample. For comparisons of the priority dimension by face location (above/below fixation), see Supplementary Methods.

In addition to the social-trait analyses described above, BTs for each face were correlated with the scores for the previously published two principal component (PC) of judgements of social traits in faces[7], that is, the two summary dimensions of social judgements. A significant correlation was found between BTs and power/dominance scores ($r = -0.39$, $P < 0.001$), but not with the valence scores ($r = -0.06$, $P = 0.27$). See Supplementary Methods and Fig. 2 for a complementary dimension-coefficients analysis.

Experiment 3. *Participants.* A total of 47 Hebrew University students (41 women, age: $M = 22.87$, s.d. = 2.66) participated for course credit ($n = 10$) or payment of 30 NIS. Forty participants reported normal vision, four had undergone laser eye surgery and three wore contact lenses.

Materials. Using FaceGen 3.1, we generated faces varying only in the priority dimension extracted from the combined sample from experiments 1 and 2. Twenty-one faces were generated, varying from $z = -5$ to $z = 5$ on the dimension. For the training phase, the same 25 faces as in previous experiments were used.

Procedure. The procedure was identical to that of experiment 2, barring the exclusion of the preliminary phase and the change of stimuli in the main blocks. The 21 faces were repeated 29 times in random order, yielding a trial count similar to previous experiments: 609 trials in total.

Statistics. Three participants reported that they had lost convergence of the visual percept during the experiment. Four participants reported that they had shut one eye while performing the tasks. One participant reported blinking excessively, and one participant reported performing rapid eye movements to expedite breaking suppression. Two participants reported only looking at one side of the screen. One participant reported squinting and demonstrated knowledge about the purpose of the experiment. Data belonging to all these participants ($n = 12$) were excluded from analysis.

Data from five participants with mean accuracy below 90% were also excluded from the analysis, leaving the data of 30 participants in the analysed sample. Data from incorrect trials were excluded from analysis ($n = 434$, 2.37%). Three trials (0.02%) with BTs of less than 200 ms were excluded from the data. Data from trials in which the BT was more than 3 s.d. from the participant's mean were excluded from analysis ($n = 327$, 1.79%).

The test–retest reliability for this sample was $r = 0.31$, $P = 0.17$, and the inter-rater agreement was ICC(2,30) = 0.30. Both reliability measures are at the lower end of the range we find in the six experiments reported here.

Additional to the ANOVA analysis reported above, a linear mixed model was fit to the data to test the hypothesis that the priority dimension predicts BTs. BTs were not aggregated or normalized per subject. Hence, BTs were log-transformed so as not to violate the assumptions of linear mixed modelling. Log-transformed BTs were predicted by the priority dimension, with random by-participant intercept and priority dimension slope. A likelihood-ratio test revealed that the priority dimension is a significant predictor of BTs (linear mixed-model coefficient, $b = 0.0016$, s.e. = 0.0008, $t = 2.02$, $\chi^2 = 4.08$, $P = 0.04$).

These results, and those reported in the main text, serve not only as a validation of the priority dimension, but also demonstrate that it does not depend on the inclusion of the preliminary adaptation phase from experiments 1 and 2, as this phase was not included in this experiment.

Experiment 4. *Participants.* A total of 40 Hebrew University students (21 female, age: $M = 23.28$, s.d. = 2.92) participated for course credit ($n = 6$) or payment of 30 NIS ($n = 34$). Of these, 35 reported normal vision and five had undergone laser eye surgery.

Materials and procedure. The procedure and presentation technique were identical to those in experiment 3, with two changes. First, for the main block, the 300 randomly generated faces from experiments 1 and 2 served as stimuli after undergoing diffeomorphic scrambling. Diffeomorphic scrambling preserves low-level visual features in images, while rendering them unidentifiable[33]. Images were scrambled 32 times (a degree of scrambling reported as suitable for faces)[33]. Second, because in pre-pilot testing it emerged that the scrambled faces take considerably longer to break through suppression, the stimuli faded into 100% instead of 35% contrast during the first second of presentation.

Statistics. Two participants reported that they had shut one eye, and one reported blinking excessively while performing the tasks. One participant reported only looking at the top part of the screen. Data belonging to all these were excluded from analysis ($n=4$). Data belonging to six participants with mean accuracy below 90% were also excluded from the analysis, leaving the data of 30 participants in the analysed sample. Data from incorrect trials were excluded from analysis ($n=599$, 3.33%). Trials with BTs less than 200 ms were excluded from the data ($n=37$, 0.21%). Data from trials in which BT was more than 3 s.d. from the participant's mean were excluded from analysis ($n=381$, 2.12%). BTs were standardized per participant.

The test–retest reliability for BTs in this sample was $r=0.29$, $P<0.001$. Inter-rater agreement of BTs for the scrambled faces stimuli was ICC(2,30) = 0.40.

We computed the scrambled face reverse correlation dimension. Predicted BTs were computed from the dimension, and the correlation between predicted BTs and actual BTs in this sample was $r=0.39$, $P<0.001$, so the reverse correlation dimension for scrambled faces explains 15.21% of the variance in BTs. Additional to the partial correlation analysis reported in the main text, the partial correlation between predicted BTs from experiment 1 and actual BTs from experiment 2, controlling for the scrambled face dimensions, is $r=0.28$, $P<0.001$, further indicating that low-level features preserved in the scrambles cannot completely explain the priority dimension.

Experiment 5. *Participants.* In total, 38 Hebrew University students participated in this experiment (31 women, age: $M=23.15$, s.d. = 2.43) in return for course credit ($n=31$) or payment ($n=7$). Of these, 34 reported normal vision, three had undergone laser eye surgery, and one used contact lenses.

Materials and procedure. The procedure was identical to that of experiment 2, except that on the main block, the 300 randomly generated faces were displayed inverted (vertical flip), thus entirely preserving low-level features but impeding holistic face processing[14].

Statistics. Seven participants reported in the debriefing form that they had shut one eye while performing the task. One participant reported looking only at one side of the screen. Data for these participants were excluded from analysis. Data belonging to one participant with mean accuracy below 90% were also excluded from analysis, leaving the data of 29 participants in the analysed sample. Experimenter error led to the termination of data collection for this experiment with 29 participants, a fact discovered only after term had ended and further data collection was not possible. Data from incorrect trials were excluded from analysis ($n=427$, 2.45%). Trials with BTs less than 200 ms were excluded from the data ($n=6$, 0.03%). Data from trials in which BT was more than 3 s.d. from the participant's mean were excluded from analysis ($n=325$, 1.87%). BTs were standardized per participant.

The test–retest reliability for BTs in this sample was $r=0.34$, $P<0.001$. Inter-rater agreement of BTs for the inverted faces was ICC(2,29) = 0.46.

The reverse correlation dimension was extracted for inverted faces. Predicted BTs were computed from the dimension. The correlation between predicted BTs and actual BTs in this sample was $r=0.53$, $P<0.001$, so the reverse correlation dimension for inverted faces explains 28.09% of the variance in BTs.

Additional to the partial correlation reported in the main text, the partial correlation between predicted BTs from experiment 1 and actual BTs from experiment 2, controlling for the inverted face dimension, is $r=0.13$, $P=0.028$. Combined with the results from experiment 4, the partial correlation between predicted BTs from experiment 1 and actual BTs from experiment 2, controlling for both the inverted and scrambled face dimensions, is $r=0.11$, $P=0.052$. Together with the correlations presented in the main text, this is an indication that while some of the priority dimension is inversion-invariant, overall low-level processing preserved in the inverted and scrambled faces cannot fully explain the priority dimension.

Experiment 6. *Participants.* In total, 54 Hebrew University students (39 women, age: $M=22.63$, s.d. = 1.32) participated for course credit ($n=31$) or payment of 40 NIS ($n=23$). Of these, 39 reported normal vision, seven had undergone laser eye surgery, and eight wore contact lenses.

Materials and procedure. Materials for this experiment were identical to those used in experiment 2. This experiment had a within-subject design. All participants first completed a conscious version of the task and then the bCFS task. The bCFS task was identical in procedure to the one used previously. On each trial in the conscious task, a face stimulus was presented binocularly, blended into a single static Mondrian at a contrast level of 35%. Thus, the face stimulus was visible and conscious from trial onset. Participants responded to the location of the face stimulus relative to fixation, as in the bCFS task. As in previous experiments, both tasks were preceded by a 25-trial training block.

Statistics. Three participants reported that they had shut one eye and one reported performing fast repeated saccades while performing the tasks. Eleven participants reported only looking at one side of the screen. Three participants had lost convergence of the visual percept during the experiment. Data belonging to all these groups were excluded. Data belonging to participants with mean accuracy below 90% were also excluded (bCFS, $n=10$; conscious task, $n=1$), leaving the data of 35 participants to be analysed for the conscious control task and 29 in the bCFS task. Data from incorrect trials were excluded from analysis (bCFS block, $n=505$, 2.90%; conscious block, $n=381$, 1.81%). Trials with BTs or conscious RTs less than 200 ms were excluded from the data (bCFS block, $n=2$, 0.01%; conscious block, $n=7$, 0.03%). Data from trials in which BTs or conscious RTs were more than 3 s.d. from the participant's mean for the task were excluded from analysis (bCFS block, $n=341$, 1.96%; conscious block, $n=334$, 1.59%).

Given that the faces were immediately visible, it is not surprising that conscious reaction times ($M=523$ ms, s.d. = 78 ms) were significantly shorter than BTs ($M=1588$ ms, s.d. = 337 ms): $t(28)=18.15$, $P<0.001$, Cohen's $d=3.37$. BTs and conscious RTs were standardized per participant. The test–retest reliability for BTs in this sample ($r=0.31$, $P<0.001$) was significantly higher than for conscious RTs ($r=0.13$, $P=0.02$): Fisher's $z=2.22$, $P=0.03$. Inter-rater agreement of BTs was ICC(2,29) = 0.34 and for conscious RTs ICC(2,35) = 0.23. Therefore, BTs seem to be a more stable differentiator of the 300 faces than conscious RTs.

Reverse correlation dimensions were extracted for BTs and conscious RTs in this sample. For both measures, the correlation between scores predicted from the dimension and actual scores was significant (BTs, $r=0.41$, $P<0.001$, dimension explaining 16.81% of variance; conscious RTs, $r=0.33$, $P<0.001$, dimension explaining 10.89% of variance).

These results, together with those reported in the main text, demonstrate the extraction of the priority dimension in an experiment in which participants did not complete the preliminary adaptation phase. We thus conclude that the priority dimension does not depend on the inclusion of such a phase.

A within-participant analysis of BTs and conscious RTs was carried out. Mean BTs and conscious RTs were computed for each face and each participant. The correlation between BTs and conscious RTs across face stimuli was then calculated for each participant. The mean correlation coefficient ($M=-0.01$, s.d. = 0.05) over participants did not differ significantly from 0 ($t(28)=0.81$, $P=0.42$). This is further evidence for the divergence of the priority dimension and the conscious RT dimension.

In addition to the social-trait comparative analysis reported in the main text, a correlation analysis revealed that RTs predicted from the conscious RT dimension are positively correlated with valence scores ($r=0.25$, $P<0.001$), while predicted BTs did not correlate with valence scores ($r=0.09$, $P=0.11$). This difference is significant (Williams's $t(297)=2.48$, $P=0.01$).

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. The R code for data processing and analysis is publicly available on the lab website of R.R.H. (http://labconscious.huji.ac.il/?p=855). The MATLAB code used to run the experiments is available from Y.A.

Data availability. All data that support the findings of this study are publicly available from http://labconscious.huji.ac.il/?p=855. The stimuli used in this study are publicly available on the lab website of A.T. (http://tlab.princeton.edu/databases/randomfaces/).

Received: 4 June 2017; Accepted: 20 November 2017;
Published online: 18 December 2017

References

1. Baars, B. J. In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *J. Conscious. Stud.* **4**, 292–309 (1997).
2. Erdelyi, M. H. A new look at the new look: perceptual defense and vigilance. *Psychol. Rev.* **81**, 1–25 (1974).
3. Mack, A. & Rock, I. *Inattentional Blindness* 33 (MIT Press, Cambridge, MA, 1998).
4. Simons, D. J. & Rensink, R. A. Change blindness: past, present, and future. *Trends Cogn. Sci.* **9**, 16–20 (2005).
5. Axelrod, V., Bar, M. & Rees, G. Exploring the unconscious using faces. *Trends Cogn. Sci.* **19**, 35–45 (2015).
6. Jack, R. E. & Schyns, P. G. Toward a social psychophysics of face communication. *Annu. Rev. Psychol.* **68**, 269–297 (2017).
7. Oosterhof, N. N. & Todorov, A. The functional basis of face evaluation. *Proc. Natl Acad. Sci. USA* **105**, 11087–11092 (2008).
8. Most, S. B., Scholl, B. J., Clifford, E. R. & Simons, D. J. What you see is what you set: sustained inattentional blindness and the capture of awareness. *Psychol. Rev.* **112**, 217–242 (2005).
9. Stoerig, P. & Cowey, A. Blindsight in man and monkey. *Brain* **120**, 535–559 (1997).
10. Gazzaniga, M. S. Forty-five years of split-brain research and still going strong. *Nat. Rev. Neurosci.* **6**, 653–659 (2005).
11. Driver, J. & Vuilleumier, P. Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition* **79**, 39–88 (2001).

12. Tsuchiya, N. & Koch, C. Continuous flash suppression reduces negative afterimages. *Nat. Neurosci.* **8**, 1096–1101 (2005).
13. Jiang, Y., Costello, P. & He, S. Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol. Sci.* **18**, 349–355 (2007).
14. Maurer, D., Grand, R. L. & Mondloch, C. J. The many faces of configural processing. *Trends Cogn. Sci.* **6**, 255–260 (2002).
15. Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**, 223–233 (2000).
16. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
17. Freiwald, W. A. & Tsao, D. Y. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**, 845–851 (2010).
18. Stein, T., Peelen, M. V. & Sterzer, P. Adults' awareness of faces follows newborns' looking preferences. *PLoS ONE* **6**, e29361 (2011).
19. Zhou, G., Zhang, L., Liu, J., Yang, J. & Qu, Z. Specificity of face processing without awareness. *Conscious. Cogn.* **19**, 408–412 (2010).
20. Yang, E., Zald, D. H. & Blake, R. Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion* **7**, 882–886 (2007).
21. Chen, Y.-C. & Yeh, S.-L. Look into my eyes and I will see you: unconscious processing of human gaze. *Conscious. Cogn.* **21**, 1703–1710 (2012).
22. Stein, T., Senju, A., Peelen, M. V. & Sterzer, P. Eye contact facilitates awareness of faces during interocular suppression. *Cognition* **119**, 307–311 (2011).
23. Gobbini, M. I. et al. Prioritized detection of personally familiar faces. *PLoS ONE* **8**, e66620 (2013).
24. Stewart, L. H. et al. Unconscious evaluation of faces on social dimensions. *J. Exp. Psychol. Gen.* **141**, 715–727 (2012).
25. Todorov, A., Dotsch, R., Wigboldus, D. H. J. & Said, C. P. Data-driven methods for modeling social perception. *Soc. Pers. Psychol. Compass* **5**, 775–791 (2011).
26. Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G. & Nakayama, K. Fitting the mind to the world: face adaptation and attractiveness after effects. *Psychol. Sci.* **14**, 558–566 (2003).
27. Dotsch, R., Hassin, R. R. & Todorov, A. Statistical learning shapes face evaluation. *Nat. Hum. Behav.* **1**, 0001 (2016).
28. Adolphs, R., Nummenmaa, L., Todorov, A. & Haxby, J. V. Data-driven approaches in the investigation of social perception. *Phil. Trans. R. Soc. Lond. B* **371**, 20150367 (2016).
29. Neri, P., Parker, A. J. & Blakemore, C. Probing the human stereoscopic system with reverse correlation. *Nature* **401**, 695–698 (1999).
30. Tadin, D., Lappin, J. & Blake, R. Fine temporal properties of center-surround interactions in motion revealed by reverse correlation. *J. Neurosci.* **26**, 2614–2622 (2006).
31. Gosselin, F. & Schyns, P. G. Superstitious perceptions reveal properties of internal representations. *Psychol. Sci.* **14**, 505–509 (2003).
32. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420–428 (1979).
33. Stojanoski, B. & Cusack, R. Time to wave good-bye to phase scrambling: creating controlled scrambled images using diffeomorphic transformations. *J. Vis.* **14**, 6 (2014).
34. Bourdev, L. & Malik, J. Poselets: body part detectors trained using 3D human pose annotations. In *2009 IEEE 12th International Conference on Computer Vision* 1365–1372 (IEEE, New York, NY, 2009).
35. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
36. Éthier-Majcher, C., Joubert, S. & Gosselin, F. Reverse correlating trustworthy faces in young and older adults. *Front. Psychol.* **4**, 592 (2013).
37. Jack, R. E., Caldara, R. & Schyns, P. G. Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *J. Exp. Psychol. Gen.* **141**, 19–25 (2012).
38. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
39. Sklar, A. Y. et al. Reading and doing arithmetic nonconsciously. *Proc. Natl Acad. Sci. USA* **109**, 19614–19619 (2012).

Acknowledgements
This research was supported by the United States–Israel Binational Science Foundation (grant no. 2013417 to R.R.H. and A.T.). The funders had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

Author contributions
Y.A., A.Y.S. R.D. A.T. and R.R.H. developed the main ideas of this research programme. Y.A. and A.Y.S. programmed the experiments. Y.A. ran the experiments and analysed data. R.D. contributed analysis methods and scripts. Y.A. and R.R.H. wrote the manuscript, and R.D., A.Y.S. and T.D. provided extensive feedback on the writing.

Competing interests
The authors declare no competing financial interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41562-017-0266-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.A. or R.R.H.

nature research

Corresponding author(s): Yaniv Abir

☐ Initial submission ☒ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

For Experiment 1 we pre-specified a sample size of at least 30 participants on the basis of previous studies applying reverse-correlation to ratings of faces. Based on the participant inclusion criteria adopted in Experiment 1, in all following experiments we pre-specified a sample size of 30 participants after exclusion of participants by their questionnaire responses and overall performance. Exclusion rates were calculated at the end of each experimentation day, sometimes resulting in small variations in the number of participants.
As the dimension extraction procedure we use does not yield itself to traditional power analysis, our a-priori decision to collect data from 30 participants was retained. Nonetheless, using a jackknife approach, we computed the expected power for our sample sizes - see Supplementary Methods section.

2. Data exclusions

Describe any data exclusions.

Outlier removal procedure were based on previous bCFS studies. Based on debriefing questionnaires, participants that had reported applying special strategies during the bCFS task (such as closing one eye, looking only at one side of the screen, or blinking excessively) were excluded from analysis. Following convention, participants with accuracy rates of less than 90% were also excluded from analysis. For the remaining data, trials with RTs of less than 200ms or that have RTs removed from the participant's mean by more than 3 SD were removed from analysis. RT analysis was performed on correct responses only.
Exact numbers of excluded participants and trials are reported for each experiment in the Methods section.

3. Replication

Describe whether the experimental findings were reliably reproduced.

In Experiment 2 we replicate our main finding of the priority-dimension, which was first identified in Experiment 1. One of the conditions in Experiment 6 replicates our main finding once again.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

In all of our experiments comparisons are within participant.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

In all of our experiments comparisons are within participant.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▸ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Analyses for all experiments were carried out using the R statistical environment 3.3.2, with the reshape, plyr, ez, psych, ggm, bootstrap, pwr and lme4 packages.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▸ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All materials are available upon request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	NA.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	NA
b. Describe the method of cell line authentication used.	NA
c. Report whether the cell lines were tested for mycoplasma contamination.	NA
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	NA

▸ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	NA

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

All of our participants were Hebrew University students, participating in return for course credit or payment of 30 NIS (about 8.5$). Mean and SD of age are given in the methods section for each experiment, as is the number of women and men participants, and the number of participants receiving credit rather than.