Evaluating Faces on Social Dimensions

 Alexander Todorov
  Princeton University

 Ghent, May 23th, 2008
A puzzle

• Superficial judgments from faces that are not necessarily accurate
  – Predict a variety of social outcomes
    • Sentencing decisions (Blair et al., 2004; Eberhardt et al., 2006; Zebrowitz & McDonald, 1991)
    • Electoral success (Ballew & Todorov, 2007; Little et al., 2007; Todorov et al., 2007)
  – Are highly efficient
    • Formed after minimal time exposure (Bar et al., 2006; Willis & Todorov, 2006)
    • Automatic categorization on trait dimensions (Engell, Haxby, & Todorov, 2007)

• What is the functional basis of these judgments?
Overview and acknowledgments

- Efficiency of judgments and the role of the amygdala
- A bottom-up driven dimensional approach
- Revisiting the role of the amygdala I
- Computer modeling of face variations on social dimensions
- Behavioral tests of the model
- Revisiting the role of the amygdala II

Andy Engell, Anesu Mandisodza, Amir Goren, Brad Duchaine, Chas Ballew, Chris Olivola, Chris Said, Crystal Hall, Janine Willis, Jim Haxby Manish Pakrashi, Nick Oosterhof, Sara Verosky, Sean Baron, Richard Lopez, Valerie Loehr
National Science Foundation
Face trustworthiness and the amygdala

- Patients with bilateral amygdala damage judge untrustworthy faces as trustworthy (Adolphs, Tranel, & Damasio, 1996)

- The amygdala seems to track to perceived trustworthiness of faces (Winston et al., 2002)
Speed of inferences from faces

Is this person trustworthy?

An ostensibly memory task in the scanner

1 Block = 53s

Each face presented for 1s
Target mean ISI = 3.5s
Minimum ISI = 1.5s
6 Blocks per run
2 runs
But what are we measuring?
A data-driven approach

• Reducing trait judgments to a simple two-dimensional space

• Modeling how faces vary on the dimensions of face evaluation

• Using the models, finding out the features that produce specific judgments

Oosterhof & Todorov (Under review, a).
Identifying trait dimensions

Spontaneous unconstrained person descriptions $\approx 1,100$

Reduction to 14 trait attributes

Faces rated by groups of raters on each trait ($\alpha \geq .90$)
A two-dimensional solution

• Two principal components that account for 83% of the variance of trait judgments
  – Evaluation
  – Dominance

• Consistent with other dimensional models (cf., Osgood, Susi, & Tannenbaum, 1957)
  – Wiggins’ model of inter-personal perception
  – Fiske’s model of inter-group perception
Factor loadings of trait judgments

- Trustworthy
- Emotionally stable
- Sociable
- Responsible
- Caring
- Attractive
- Intelligent
- Confident
- Dominant
- Unhappy
- Aggressive
- Mean
- Weird

Correlation with latent evaluation factor
Re-evaluating the role of the amygdala I

• Is the amygdala really responding to trustworthiness?

• Extracting the mean signal change in face-responsive voxels in the amygdala for each face

• Testing the valence hypothesis

Todorov & Engell (under review).
(a) Brain imaging showing activity in the amygdala.

(b) Heatmap illustrating the correlation between amygdala activity and emotional valence. The labels on the right indicate significant correlations at the p < .05 level.

(c) Scatterplot showing the relationship between right amygdala activity and valence ratings of faces.

(d) Scatterplot showing the relationship between left amygdala activity and valence ratings of faces.
A data-driven approach

• Reducing trait judgments to a simple two-dimensional space

• Modeling how faces vary on the dimensions of face evaluation

• Using the models, finding out the features that produce specific judgments

Oosterhof & Todorov (Under review, a).
Model for face representation: Facegen

- Computerized 3D face model
- Based on Principal Components Analysis
  - Data driven method
    - Aimed at representing the most plausible faces
    - Deviation from ‘average’ face
    - 50D for symmetric shape
- Advantages
  - Generate unlimited number of random faces
  - Precise control over stimuli
- No a priori assumptions about importance of facial features
Face shape for specific faces is represented as a function of the average face and the differences from this face for all vertices.
Building a trait dimension

<table>
<thead>
<tr>
<th>Trait judgment</th>
<th>Principal components</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.20</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>300</td>
<td>+2.13</td>
</tr>
<tr>
<td>.</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
PC1 & PC2 & \ldots & PC50 \\
.25 & . & . & .12 \\
-.30 & . & . & -.10 \\
. & . & . & . \\
. & . & . & . \\
. & . & . & . \\
1.0 & . & . & -.25 \\
\end{bmatrix}
\]

= f
Trustworthiness

Dominance
Model validation

A

B

Trustworthiness judgments
Trustworthiness Dimension (SD)

Dominance judgments
Dominance Dimension (SD)
A data-driven approach

• Reducing trait judgments to a simple two-dimensional space

• Modeling how faces vary on the dimensions of face evaluation

• Using the models, finding out the features that produce specific judgments

Oosterhof & Todorov (Under review, a).
Neutral or expressing one of the basic emotions

angry
disgusted
sad
fearful
neutral
surprised
happy

Trustworthiness (SD)  Dominance (SD)

proportion
0.71

chance
0
trustworthiness - untrustworthiness
submissiveness - dominance
trustworthiness - untrustworthiness
submissiveness - dominance
Holistic data driven model

• The input to the model is completely unbiased with respect to emotions
  – Emotionally neutral faces
• Exploits the variation in trustworthiness judgments
• Expressions of anger and happiness fall out of this variation
The functional basis of face evaluation

• Faces are evaluated on two fundamental dimensions

• These evaluations originate in adaptive mechanisms
  – Evaluations = Inferring intentions to harm
    • features resembling expressions signaling approach/ avoidance behavior
  – Dominance = Inferring the ability to cause harm
    • features signaling physical strength

• Other social judgments can be represented as a function of these two dimensions
The source of valence evaluation

• Overgeneralization of mechanisms for recognition of emotional expressions? (Secord, 1958; Knutson, 1996)
• Consistent with computer modeling findings
• Three additional types of evidence
  – Correlations between judgments of trustworthiness and judgments of anger and happiness (Todorov & Duchaine, in press)
    • Similar effects for affiliation (Monterare & Dobish, 2002)
  – Dynamic stimuli
  – Behavioral adaptation
Perceiving changes in expression

Same face  Congruent morph  Incongruent morph
Oosterhof & Todorov (Under review, b).
Re-evaluating the role of the amygdala II

• Is the amygdala response to the valence of faces linear?

• Most fMRI studies of expression perception find increased response to happy faces

• If sufficient range of trustworthiness, possible non-linear response to face trustworthiness
Said, Baron, & Todorov (in press). JCN.
Todorov, Baron, & Oosterhof (in press). SCAN.
The role of the amygdala in valence evaluation of faces

• In implicit contexts, a relatively linear response
  – At least within reasonable range of valence

• Possibly guiding appropriate approach/avoidance behavior (D. Amaral, 2002)

• In explicit contexts of evaluation, possibly non-linear response
  – And possibly within extended range of valence
Methodological implications: The value of formal representational face models

• Making the invisible visible
• Precise control of stimuli
• Objective scaling of differences on social dimensions
  – Modeling of behavioral judgments
  – Range of face trustworthiness
    • comparisons across studies
  – Theory validation approaches
Possibly resolving the puzzle

• Evaluating emotionally “neutral” faces is an overgeneralization of functionally adaptive inferences
  – Harmful intentions
  – The ability to cause harm

• This can account for the efficiency and subjectively compelling character of these processes