

# Experience-dependent coding of facial expression in superior temporal sulcus

Nicholas Furl<sup>\*†</sup>, Nicola J. van Rijsbergen<sup>‡</sup>, Alessandro Treves<sup>‡</sup>, Karl J. Friston<sup>\*</sup>, and Raymond J. Dolan<sup>\*</sup>

<sup>\*</sup>Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom; and <sup>‡</sup>Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, via Beirut 2/4, 34104 Trieste, Italy

Edited by David Mumford, Brown University, Providence, RI, and approved July 6, 2007 (received for review March 19, 2007)

**Sensory information from the external world is inherently ambiguous, necessitating prior experience as a constraint on perception. Prolonged experience (adaptation) induces perception of ambiguous morph faces as a category different from the adapted category, suggesting sensitivity in underlying neural codes to differences between input and recent experience. Using magnetoencephalography, we investigated the neural dynamics of such experience-dependent visual coding by focusing on the timing of responses to morphs after facial expression adaptation. We show that evoked fields arising from the superior temporal sulcus (STS) reflect the degree to which a morph and adapted expression deviate. Furthermore, adaptation effects within STS predict the magnitude of behavioral aftereffects. These findings show that the STS codes expressions relative to recent experience rather than absolutely and may bias perception of expressions. One potential neural mechanism for the late timing of both effects appeals to hierarchical models that ascribe a central role to backward connections in mediating predictive codes.**

hierarchical bayes | magnetoencephalography | predictive coding | top-down processing | visual evoked fields

Despite ambiguity about what causes retinal images, human vision overcomes uncertainty and ambiguity to parse stimuli into well defined categories, such as identity and expression in faces, necessary to support complex social interactions. A compelling demonstration comes from adaptation aftereffects, where prolonged exposure to a stimulus (adaptation) biases subjects to perceive stimuli as dissimilar to adapting stimuli. Thus, visual perception is not veridical but uses previous experience as a referent, often described as a norm (1–4). High-level aftereffects have been shown for face categories, including identities (2, 4), races, genders, and expressions (5). Also, recent studies suggest that the visual system emphasizes the deviation between individual and average (norm) faces (3, 6). Face adaptation therefore provides a powerful paradigm for investigating the role of experience in visual coding.

A number of neural mechanisms might mediate these perceptual influences of experience. Relatively simple models propose that adaptation desensitizes feedforward neural pathways specialized for detecting the adapted pattern (2, 7), perhaps by fatigue (8), which allows competing feature detectors to influence perception. Adaptation can also be viewed from the standpoint of hierarchical predictive coding models (8–12), which characterize visual representations as a “prediction error,” reflecting the deviation between bottom-up stimulus patterns and top-down expectations tuned by previous experience. When environmental contingencies change, as exemplified by adaptation, these predictions are dynamically recalibrated (i.e., sensory and perceptual learning). An important feature of predictive coding models is that predictive codes are mediated by backward connections from more advanced areas (8–13). Both types of model predict that perceptual codes for facial categories will be experience-dependent, in that neural populations will emphasize deviations from recent experience.

Employing magnetoencephalography (MEG) we asked, using precise temporal resolution, how and when adaptation to facial expression categories modulates neurophysiological responses to ambiguous morph faces intermediate between adapting expressions. We identified expression codes that were dependent on adapted experience by testing for evoked neuromagnetic signals parametrically encoding deviations between adapted and morph expressions. In addition, we localized neuromagnetic signals correlated with the behavioral bias induced by face adaptation. Lastly, we reconstructed the sources of these effects to identify the anatomic structures contributing to integration of recent adapting experience during perception.

## Results

We used MEG to measure neuromagnetic responses evoked by linearly interpolated morph faces varying in fear expression (10%, 30%, 50%, 70%, and 90%) after adaptation to scrambled (control), neutral, and fearful facial expressions (Fig. 1*A*). For each adaptation condition, 12 subjects were repeatedly exposed to an adapting expression and then performed a sequence of categorization trials, each comprising a top-up adapting stimulus followed by a morph that was categorized as fearful or neutral. Performance in this paradigm (Fig. 1*B*) replicated previous findings (5) that adaptation shifts the category boundary (the transition between neutral and fearful categorizations, somewhere midway through the continuum), reflecting bias to perceive the nonadapted category.

The morphs evoked neuromagnetic fields, including an M100 and M170, associated with face processing (14) as well as a later sustained field (Fig. 1*C* and *D*). We tested for adaptation effects on evoked fields in sensor space by generating, at every post-stimulus time point, two-dimensional interpolated spatial maps of the sensor data (Fig. 1*D*) and then stacking them in time to form three-dimensional spatiotemporal volumes amenable to general linear model statistics with family-wise error correction based on random-field theory (15). This analysis furnishes spatiotemporal statistical parametric maps in sensor space of the *F* statistic for every point in space and time, testing for experimental effects over subjects. We then reconstructed the anatomic sources for each of the subjects, in each condition, by using distributed minimum-norm solutions (16). These reconstructions were in the standardized space of Talairach and Tournoux, enabling us to test hypotheses at the between-subject level in source space, with Statistical Parametric Mapping (Wellcome Department of Imaging Neuroscience, London) in the conventional way.

Author contributions: N.F., N.J.v.R., A.T., and R.J.D. designed research; N.F. and N.J.v.R. performed research; K.J.F. contributed new reagents/analytic tools; N.F. analyzed data; and N.F., A.T., and R.J.D. wrote the paper.

The authors declare no conflict of interest.

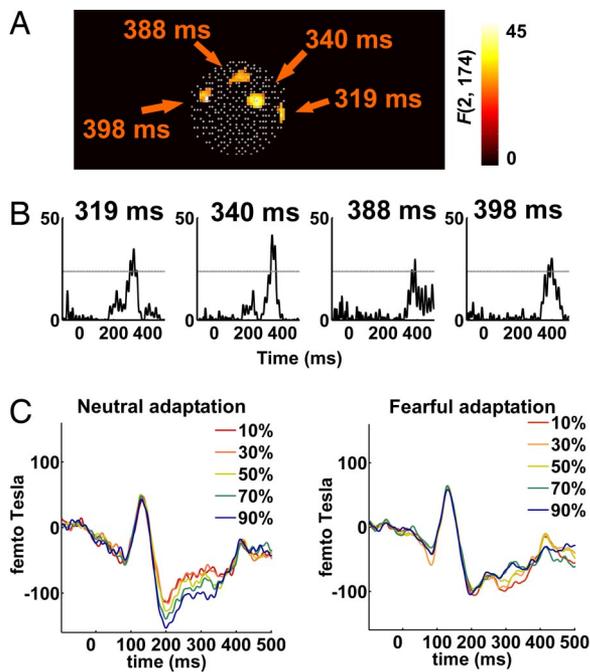
This article is a PNAS Direct Submission.

Abbreviations: MEG, magnetoencephalography; SPM, statistical parametric map; STS, superior temporal sulcus.

<sup>†</sup>To whom correspondence should be addressed. E-mail: n.furl@fil.ion.ucl.ac.uk.

© 2007 by The National Academy of Sciences of the USA



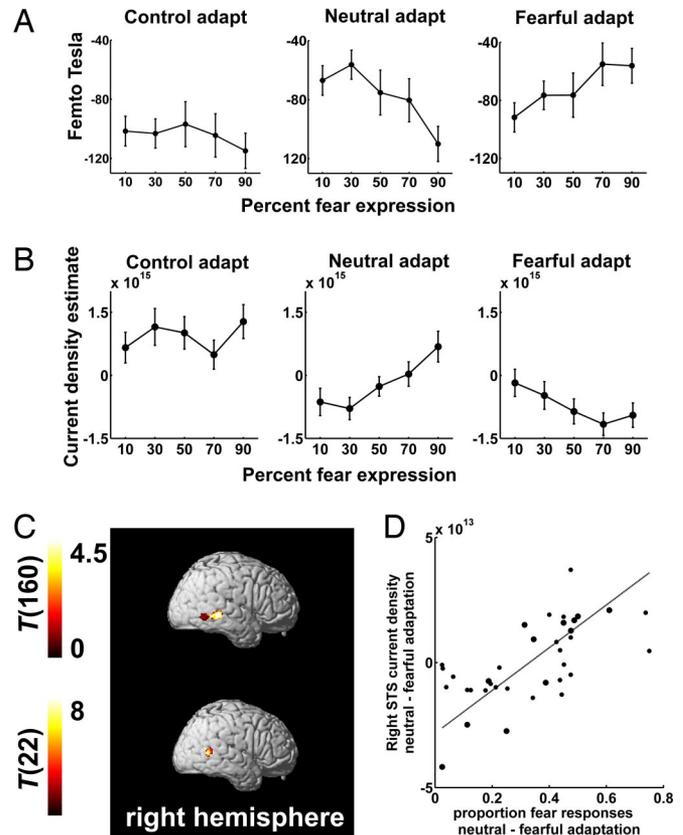


**Fig. 3.** Interactions between adapted and morph expressions. (A) SPMs of the  $F$  statistic, thresholded at  $P < 0.05$  (family-wise error-corrected). Each cluster is labeled by the time of maximal effect. (B) Time course of  $F$  statistic at the maximal voxel within each cluster labeled by the time of peak effect, all between 300 and 400 ms after stimulus. The horizontal dashed line represents the  $P < 0.05$  (family-wise error-corrected) significance threshold. (C) Waveforms manifested at the right lateral temporal sensor showing a peak effect at 319 ms, plotted for neutral adaptation (Left) and fearful adaptation (Right).

pressions are coded relative to recent experience rather than absolutely. We also tested for evidence of absolute coding, which predicts a parametric effect of morph expression independent of experience. However, no effects of morph expression were observed at any time point when collapsing over adaptation conditions. Thus, the evidence favors an account where morph expressions are coded by using recent experience as a reference rather than according to absolute emotional intensity.

After reconstructing sources between 300 and 400 ms, we tested for interactions between adapting expression and the degree of morphing by contrasting correlations with morph expression after neutral versus fearful adaptation. We found significant effects ( $P = 0.01$ , corrected) within right superior temporal sulcus (STS) (peak: 60,  $-36$ ,  $-2$  mm; Brodmann area 21), a region implicated in expression processing (20), which adapts to repeated expressions (21). The pattern of source estimates in the right STS (Fig. 4B) reproduces the pattern of effects observed in sensor space (Fig. 4A). In the left hemisphere, we observed a smaller, more posterior cluster in middle temporal gyrus (peak:  $-56$ ,  $-68$ , 6 mm; Brodmann area 37;  $P = 0.03$ ; corrected).

Finally, we examined how STS activity during this late period (300–400 ms) related to subjects' actual perception. We correlated adaptation effects on explicit fear categorizations (neutral versus fearful adaptation) with effects in source space (neutral versus fearful adaptation), collapsing over 30–70% morphs, where there was maximal variation in aftereffect size. In posterior STS, close to the temporal parietal junction (Fig. 4C: 44,  $-50$ , 8 mm; Brodmann area 39), larger behavioral aftereffects were significantly ( $P < 0.0001$ ) predicted by the magnitude of adaptation effects on source estimates, across all 12 subjects (Fig. 4D).



**Fig. 4.** Adaptation effects arising 300 ms after stimulus. (A) Mean  $\pm$  SE of neuromagnetic sensor space responses to morphs at 319 ms in maximal voxel for that cluster. Because these means are extracted from a negatively deflected component, more negative values represent more intense responses (Fig. 3C). (B) Mean  $\pm$  SE of estimated current density (mean-corrected) in right STS voxel showing differential effects of adapted expression on correlations with morph level. In this case, more positive values represent more intense responses. Note that source space estimates reproduce pattern observed in sensor space. (C) SPMs of the  $T$  statistic thresholded at  $P < 0.05$  (family-wise error-corrected) and rendered on cortical surface showing effects on anatomic sources reconstructed between 300 and 400 ms, including interactions between adapted and morph expressions in the right STS (Upper) and correlations between adaptation effects in the right posterior STS and the size of the behavioral aftereffect (Lower). (D) Scatterplot and regression line showing correlation between adaptation effects in source space (neutral versus fear adaptation) and the size of the behavioral aftereffect in individual subjects (fear responses, neutral versus fear adaptation) by using data from 30–70% morph levels.

## Discussion

We investigated the neural dynamics underlying how recent perceptual experience (in the form of adaptation) is integrated during the perception of ambiguous face expressions. We replicated the finding (18) that face adaptation (relative to nonface control adaptation) attenuates the magnitude of the relatively early M170 response. Interestingly, we also found that fearful adaptation gives rise to a more attenuated M170 than neutral adaptation. This effect of expression adaptation could occur for numerous reasons, including increased attention to emotional information during the adaptation period. Another possibility is that fearful adaptation may affect additional pathways specialized for emotion processing. For example, a putative subcortical pathway has been proposed that is sensitive to fearful expressions and that may modulate cortical responses to faces (22). When the source was reconstructed, these M170 effects were associated with right fusiform gyrus, an area known to adapt to faces and implicated in domain specific face processing (19).

Although the M170 showed sensitivity to adapted expression, it did not show differential responses to morph expression in any of the adaptation conditions. Therefore, the M170 is unlikely to reflect neural activity that codes the expression of the input stimulus.

Differential effects of morph expression were instead manifested by a later MEG component. During a sustained field between 300 and 400 ms, current density in the mid-STS was parametrically enhanced by the difference between adapted and morph expressions. During this same period, adaptation effects in adjacent posterior STS correlated with the size of the aftereffect (the difference in fear perception induced by adaptation). These STS MEG responses reflect the difference between morph and adapted expressions and predict the size of subjects' adaptation-induced shift in perception of morphs away from the adapted expression. Importantly, these results confirm proposals that STS is specialized for representing facial actions such as expressions (20) and adapts to repetitions of facial actions such as expression (21), gaze (23), and pose (24). These results extend these proposals by suggesting that expression coding in the STS is experience-dependent, emphasizing differences between stimulus expressions and recent experience.

These findings are also compatible with norm-based coding models, which characterize facial stimuli as vectors in a multi-dimensional feature space centered on a prototypic norm (25). Faces are coded in this "face space" as the set of deviations from a retuneable statistical reference. During adaptation, this norm shifts toward the adapting stimulus, exaggerating the deviation from the norm of vectors pointing away from the direction of adaptation in face space (2, 4). Although variants of norm-based representation have been invoked to explain a variety of low-level aftereffects since Gibson (1), evidence has more recently accumulated in the domain of face perception, including recent findings that neural populations code for the deviation from averaged norm faces (3, 6).

Functional accounts such as norm-based coding, which characterize stimuli as coded with reference to experience, effectively describe a form of novelty detection. That is, experience-dependent visual codes emphasize information that is novel, compared with recent experience. Concomitantly, the perceptual shift away from the adapted expression may be described as a bias toward novel percepts, further suggesting a role for novelty-sensitive mechanisms. An interesting question, therefore, concerns the relationship between the experience-dependent responses observed here to those of previously studied novelty-sensitive responses. For example, evoked potential components such as the P300 (11) and P3B (26) respond to stimuli that deviate from recent experience and bear a late time course similar to that of the experience-dependent signal observed in the STS. The shared timing and experience-dependent properties of these effects raise an interesting hypothesis that they may also share some underlying neural mechanisms (11).

What neural mechanisms might manifest these effects? Proposals for neural mechanisms that might implement norm-based coding and perceptual aftereffects often emphasize specialized category-detecting feedforward pathways that become desensitized to adapted stimuli during adaptation, perhaps by fatigue (2, 13, 27). This process enhances visual responses to novel or nonadapted stimuli. Alternatively, categorization dynamics and the ensuing category-specific adaptation might depend predominantly on local recurrent interactions.<sup>§</sup> A further intriguing possibility relates to hierarchical generative models (8, 9, 11, 12), which emphasize backward connections that instantiate top-down, higher-level experience-based representations. These

backward connections would transmit suppressive predictive signals that induce responses reflecting prediction error. Adaptation recalibrates the prediction, such that the adapted stimulus pattern begins to evoke reduced prediction error responses. Likewise, "recovery from adaptation" constitutes heightened prediction error to novel (nonadapted) stimulus patterns. Feedforward, recurrent, and feedback mechanisms all hypothesize visual codes that are novelty-detecting and dependent on recent experience and hence predict the pattern of effects observed herein.

Although the results reported here do not distinguish these models definitively, the late timing of these effects poses some difficulty for models that posit attenuation of feedforward processing and probably also for recurrent models. If adaptation modulated feedforward sensitivity, then one would predict that experience-dependent signaling would appear very early in the time course (13). The earliest effects of adaptation appear during the M170, which has been characterized as reflecting rapid, preattentive feedforward processing (17). Nevertheless, the latency of the M170 is sufficiently long that we cannot rule out an influence of backward connections. Experience-dependent STS responses to morph expressions appeared even later, past 300 ms. A feedforward effect might be expected to manifest earlier in the time course. These results suggest predictive feedback models as an intriguing possibility, which might be confirmed by further research. It is important, for example, to establish the source of the predictive feedback. One potential candidate for the origin of this feedback is the ventromedial frontal cortex, given recent evidence that this region provides top-down feedback during perception of objects (28) and faces (29). Another possibility is that the backward connectivity is not so long-range, and the source may reside in or near the STS itself.

This work examined the effects of experience (adaptation) on behavioral and neuromagnetic responses to faces bearing ambiguous expressions. M170 attenuation to repeated faces was associated with current density in right fusiform gyrus. After 300 ms, STS current density emphasized the contrast between morph expression and recent experience, and adaptation in STS also predicted the magnitude of the perceptual aftereffects. These results contribute to our understanding of visual specialization in fusiform gyrus and STS, they complement notions of norm-based coding and novelty detection, and they pose important questions about the respective roles of feedforward and backward influences on visual coding.

## Materials and Methods

**Subjects, Stimuli, and Task.** Twelve subjects participated. Informed consent was obtained in accordance with procedures approved by The Joint Ethics Committee of The National Hospital for Neurology and Neurosurgery and The Institute of Neurology, London. Stimuli consisted of images of eight facial identities (half were female), selected from the KDEF database (30). We constructed continua of morph images (FantaMorph; Abrosoft, Beijing, China) spanning fearful and neutral expressions for each of eight identities (Fig. 1A) that were converted to gray scale, placed within a black oval frame, and equated for mean luminance. Nonface control stimuli were generated by reconstructing the face images using original spectra but randomly permuted phases, applying the same image mask as the morph images, and equating the images for mean luminance. Subjects categorized morphs as fearful or neutral in control, neutral, and fearful expression adaptation conditions. The control adaptation block always preceded neutral and fearful expression adaptation blocks (which were counterbalanced). For each adaptation condition, subjects first viewed 20 presentations of the adapting stimulus each for 500 ms followed by a 200-ms fixation interval. During the following categorization trials, subjects viewed the adapting image for 300 ms (i.e., a top-up, to refresh adaptation

<sup>§</sup>Akrami, A., Liu, Y., Treves, A., Jagadeesh, B. 36th Annual Meeting of the Society for Neuroscience Abstracts, October 14–18, 2006, Atlanta, GA, p. 504.9.

throughout the categorization task) followed by fixation for 200 ms, and then a morph face for 300 ms with the same identity as the top-up. Categorization trials were separated by a 1-s inter-trial interval. There were 80 presentations of each morph level in each adaptation condition.

**Data Acquisition and Analysis.** MEG recordings were made in a magnetically shielded room by using a 275-channel CTF system with SQUID-based axial gradiometers (VSM MedTech Ltd., Couquitlam, BC, Canada). Neuromagnetic signals were digitized continuously at a sampling rate of 480 Hz. Participants made behavioral responses with an MEG-compatible response pad, held in the right hand, and eye blinks were monitored. Data were analyzed with SPM5 (Wellcome Department of Imaging Neuroscience, London) and MATLAB. The raw continuous time series for each subject was subjected to a Butterworth band pass filter at 0.5–50 Hz. Baseline-corrected epochs were extracted from the data beginning 100 ms before stimulus onset and ending 500 ms after stimulus onset. Before averaging, all trials were removed for which the i View X software (SensoMotoric Instruments, Needham, MA) registered a blink. Averaged sensor data were converted to 3D spatiotemporal volumes by “stack-

ing” 2D linearly interpolated sensor images in peristimulus time. These spatiotemporal volumes were submitted to mass univariate general linear models by using Gaussian random field theory to control the family-wise error rate (15). Distributed reconstructions of the current density for 7,204 dipolar sources embedded in a template cortical mesh (coregistered with sensor data by using three fiducial markers) and oriented perpendicularly to the cortical surface were estimated by using parametric empirical Bayes (16) with a minimum-norm prior. Reconstructions for each subject in each condition were interpolated into Talairach and Tournoux standardized template space and then analyzed by using conventional statistical parametric mapping procedures. This analysis involved testing for main effects and interactions with mass-univariate general linear models with Gaussian random-field theory to control the family-wise error rate.

We thank Dean Mobbs for supplying the morph stimuli and Jean Daunizeau, Rik Henson, James Kilner, and Jeremie Mattout for helpful discussion and advice. This work was supported by Human Frontier Science Program Grant RGP0047/2004-C (to A.T. and R.J.D.) and Wellcome Trust Program Grant 078863 (to R.J.D.).

- Gibson JJ (1937) *Psychol Rev* 44:222–224.
- Rhodes G, Jeffery L (2006) *Vision Res* 46:2977–2987.
- Leopold DA, Bondar IV, Giese MA (2006) *Nature* 442:572–575.
- Leopold DA, O’Toole AJ, Vetter AJ, Blanz V (2001) *Nat Neurosci* 4:89–94.
- Webster MA, Kaping D, Mizokami Y, Duhamel P (2004) *Nature* 428:557–561.
- Loffler G, Yourganov G, Wilkinson F, Wilson HR (2005) *Nat Neurosci* 8:1386–1390.
- Köhler W, Wallach H (1944) *Proc Am Philos Soc* 888:269–357.
- Hosoya T, Baccus SA, Meister M (2005) *Nature* 436:71–77.
- Rao RP, Ballard DH (1999) *Nat Neurosci* 2:79–87.
- Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) *Proc Natl Acad Sci USA* 99:15164–15169.
- Friston K (2005) *Philos Trans R Soc London B* 360:815–836.
- Yuille A, Kersten D (2006) *Trends Cogn Sci* 10:301–308.
- Grill-Spector K, Henson R, Martin A (2006) *Trends Cogn Sci* 10:14–23.
- Liu J, Harris A, Kanwisher N (2002) *Nat Neurosci* 5:910–916.
- Kiebel SJ, Kilner J, Friston KJ (2007) in *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, eds Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD (Academic, London), pp 211–231.
- Mattout J, Phillips C, Daunizeau J, Friston KJ (2007) in *Statistical Parametric Mapping: The Analysis of Functional Brain Analysis*, eds Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD (Academic, London), pp 367–376.
- Furey ML, Tanskanen T, Beauchamp MS, Avikainen S, Uutela K, Hari R, Haxby JV (2006) *Proc Natl Acad Sci USA* 103:1065–1070.
- Harris A, Nakayama K (2006) *Cereb Cortex* 17:63–70.
- Kanwisher N, Yovel G (2007) *Philos Trans R Soc London B* 361:2109–2128.
- Calder AJ, Young AW (2005) *Nat Rev Neurosci* 6:641–651.
- Winston JS, Henson RN, Fine-Goulden MR, Dolan RJ (2004) *J Neurophysiol* 92:1830–1839.
- LeDoux JE (1996) *The Emotional Brain* (Simon and Schuster, New York).
- Calder AJ, Beaver JD, Winston JS, Dolan RJ, Jenkins R, Eger E, Henson RNA (2007) *Curr Biol* 17:20–25.
- Fang F, Murray SO, He S (2007) *Cereb Cortex* 17:1402–1414.
- Valentine T (1991) *Q J Exp Psychol A* 43:161–204.
- Campanella S, Quinet P, Bruyer R, Crommelinck M, Guerit JM (2002) *J Cogn Neurosci* 14:210–227.
- Menghini F, van Rijsbergen N, Treves A (2007) *Neurocomputing* 70:2000–2004.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, et al. (2006) *Proc Natl Acad Sci USA* 103:449–454.
- Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006) *Science* 314:1311–1314.
- Lundqvist D, Litton A, Öhman A (1998) *Karolinska Directed Emotional Faces* (Karolinska Institute, Stockholm, Sweden).