doi:10.1093/scan/nsn045 SCAN (2008) 3, 299–302

In this Issue

Trust and valence processing in the amygdala*

In this issue of *SCAN*, Todorov and Engell report a new study showing not only that human amygdala activation to faces is most strongly modulated by the perceived (un)trustworthiness of faces among a series of 14 personality trait dimensions (such as attractiveness, aggressiveness, intelligence, caring, and so forth), but also that such effects appear to reflect a more general response to negative valence and may arise in an 'implicit' manner while observers are engaged in a memory task (without any requirement to make explicit affective or social judgments on faces). The important conclusion drawn by these authors is that the amygdala has a key role for an automatic appraisal of the valence of unknown faces, rather than for processing other specific attributes. These findings are novel and intriguing, but also raise a number of questions about the exact meaning of valence, trust and automaticity in such appraisals.

The study of Todorov and Engell has several strengths, including an innovative methodological approach to identify the brain response to single face stimuli as a function of multiple, co-existing and correlated, psychological dimensions associated with these faces. Moreover, it also goes beyond these traditional psychological categories such as trustworthiness or attractiveness by testing for more general dimensions that may then allow a better understanding of the features used for such complex social judgments, and their corresponding neural substrates. More generally, this research also provides a valuable bridge between neuroscience studies focusing on the role of the amygdala in emotion processing and those underscoring a more general involvement in social cognition, as can be demonstrated for example by its robust response to faces with emotionally 'neutral' expressions relative to other visual object categories (Fitzgerald *et al*., 2006; Pourtois *et al*., in press). Altogether, these data clearly indicate that neutral faces are not socially or affectively neutral for the brain, and further support the view that the human amygdala function might be intimately connected with the evolution of social behavior (Amaral, 2003). Furthermore, Todorov and Engell also demonstrate that amygdala responses seem to influence concomitant activation of several cortical areas involved in face perception, suggesting that an evaluation of the trustworthiness or valence of faces might contribute to regulate attention as well as subsequent memory traces for faces.

However, what do we really learn from the notion that that amygdala may encode a general valence dimension? With respect to emotion research, such conclusion might constitute an integrative framework for the (re)interpretation of classical findings concerning the amygdala's domain of processing. Indeed, since the 1930s, the amygdala has typically been associated with a specific emotion: fear. This association was supported by evidence from animal research, brain imaging, and patient studies showing that the amygdala is important for fear learning and processing threat-related information (Phelps and LeDoux, 2005). This perspective led to the view that the amygdala is central to a 'defense system', or even a 'fear module' (Öhman and Mineka, 2001). In contrast, Todorov and Engell's interpretation is consistent with more recent results in affective neuroscience indicating that the amygdala is also involved in the processing of negative emotions other than fear (e.g. Blair *et al*., 1999; Sergerie *et al*., 2008). Furthermore, as discussed by Todorov and Engell, their finding that faces evaluated as negative do activate the amygdala does not necessarily mean that the role of the amygdala is restricted to the processing of negative stimuli. In fact, several results suggest a U-shaped amygdala response to emotional stimuli: both (intense) negative and (intense) positive (e.g. Winston *et al*., 2007). It is possible that the faces used by Todorov and Engell did not allow such a U-shaped response because more intense positive faces might have been necessary. Indeed, although many results now indicate that the amygdala is sensitive to positive stimuli (with an effect size being even larger than for negative stimuli in a meta-analysis, see Sergerie *et al*., 2008), the conditions for eliciting such responses still need to be clarified and may partly depend on the current task demands or context (e.g. Cunningham, 2008). Further research would be particularly valuable as there is no agreement in the literature on the underlying processes that subserve valence (Collombetti, 2005; Brosch and Moors, in press). Indeed, while the classical view considers valence as a single bipolar dimension (ranging from negative to positive, with neutrality in between), other models have proposed a composite of two dimensions (Russell and Carroll, 1999), including bivariate models of valence in which positive and negative dimensions may vary independently (Cacioppo *et al*., 2004). Such models are

*Commentary on Todorov and Engell.

interesting because they allow multiple, and possibly opposite, values to be assigned simultaneously to the same object (or face) when the latter conveys different types of information simultaneously; for example, a *very attractive but untrustworthy person* could be evaluated as *both* positive and negative. Whether these two sources of valence information would be integrated in the amygdala is unknown. Therefore, although the involvement of a general valence dimension proposed by Engell and Todorov provides a convenient summary for their result, this is unlikely to constitute a complete account for the pattern of amygdala responses to faces unless the exact mechanisms underlying the appraisal of value are more clearly specified.

Noteworthy, the classic dimensional models of emotion (such as the circumplex model) typically consider a second dimension, independent of valence, as being critical for affective processes; namely, arousal (e.g. Russell and Carroll, 1999). As pointed by Todorov and Engell, the valence of their faces was also likely correlated with their arousal values, and amygdala activity might be particularly sensitive to this dimension of a stimulus, irrespective of its valence (Hamman, 2003). Therefore, an alternative explanation of their results would be that the amygdala does not evaluate novel faces along a general valence dimension, but along a general arousal dimension, with the more positive arousing faces being unfortunately not directly tested in their study. However, we feel that this is very unlikely. First of all, it is important to note that the previous manipulations used to suggest that the amygdala primarily responds to arousal (e.g. Anderson *et al.*, 2003) do not truly correspond to the arousal or activation dimension originally implied by bi-dimensional models of emotion—but rather refer to the felt intensity of sensory stimuli. Note also that the 'valence by arousal' models of emotion seek to account for a particular emotional response (i.e. feeling states), rather than for the encoding of emotionally significant information (Sander *et al.*, 2005). In addition, it has been shown that the amygdala is critical for the recognition of stimuli that are low on arousal, such as sad faces (Adolphs and Tranel, 2004), whereas equally intense stimuli may differentially activate the amygdala (Whalen *et al.*, 2001), and some results indicate that neither valence nor intensity per se is coded in the amygdala (Winston *et al.*, 2005).

Nevertheless, a key role of the amygdala in terms of a general evaluative process could well explain the results of Todorov and Engell, beyond the notions of a bipolar valence dimension or arousal response. Indeed, there is increasing evidence that the amygdala plays a critical role for detecting information that is appraised as self-relevant, based on one's needs, values, goals, or concerns (Sander *et al.*, 2003). This proposal is consistent with a preferential processing and learning for fear-related stimuli in many contexts, because such stimuli are usually more relevant than others in similar conditions. Likewise, arousing stimuli are typically more relevant than non-arousing stimuli (in fact, arousal is often considered as a diagnostic response to relevant stimuli, see Sander *et al.*, 2005). More critically, this proposal may also explain why low arousal (e.g. sad faces) and so-called 'emotionally neutral' facial (e.g. Fitzgerald *et al.*, 2006; Wright and Liu, 2006) or non-facial (e.g. Ousdal *et al.*, 2008; Schaefer *et al.*, 2006) stimuli can activate the amygdala as they still have a particular affective impact on the individual. Thus, the findings of Todorov and Engell that 'emotionally neutral' faces may engage the amygdala as a function of some general value associated with perceived (un)trustworthiness is entirely consistent with an account in terms of self-relevance and points towards a general appraisal process that includes the social judgment of other persons, rather than a purely emotional process. Furthermore, this general value processed in the amygdala (and interconnected face processing regions) might encompass both the 'good' and 'bad' extremes of the traditional valence dimension.

Another surprising aspect revealed by the findings of Todorov and Engell, as well as from the pioneer study on face trustworthiness by Winston *et al.* (2002), is that such complex social appraisals may take place in a seemingly 'automatic' manner. That is, the relevant facial information appears to be extracted despite a lack of explicit task requirement. However, what can we really learn from 'automaticity' in such conditions? Todorov and Engell's results are broadly consistent with other findings that some evaluative processes may activate the amygdala without conscious control or even without conscious awareness (e.g. Öhman and Mineka, 2001; Vuilleumier, in press). Moreover, abundant evidence from patient studies and functional neuroimaging in normal individuals suggest that the amygdala may respond to emotion-eliciting information covertly, when voluntary attention is directed to another object in the display (Vuilleumier, 2005), or implicitly, when attention is focused on another non-emotional dimension of faces (see review in Vuilleumier *et al.*, 2003). However, automaticity might be a relatively ill-defined notion that encompass a number of distinct features which may or may not apply for different appraisal processes—including not only unawareness but also non-intentionality and goal-independence, lack of cognitive control, resistance to resource depletion, as well as efficiency and rapidity (Moors and De Houwer, 2006). Thus, although amygdala responses to some emotional stimuli may exhibit several of these properties, it remains unclear whether truly unconscious and implicit activation is evoked only by simple emotion signals that are conveyed by elementary facial features (Whalen *et al.*, 2004; Vuilleumier, 2005). Although some authors have suggested that complex moral judgments may also arise automatically and unconsciously in the amygdala (e.g. Moll *et al.*, 2002), it seems unlikely that such high-level social processes might be automatic in the strong sense, implying a total independency from capacity demands and current goals (Moors and De Houwer, 2006). As noted by Todorov and Engell, the amygdala involvement in trust evaluation, as well as other social or emotional judgments, might be sensitive to expectations and top-down context-dependent mechanisms (Pessoa, 2008). An alternative way to consider the level of processing and degree of automaticity in amygdala activity might be to further clarify the boundary conditions under which evaluative mechanisms may occur or may not. Thus, implicit amygdala responses such as those observed by Todorov and Engell might primarily reflect some 'default setting' rather than a hard-wired, unflexible automatic process. In other words, while the brain might be somehow prepared to process and organize sensory information along specialized 'highways' under most current situations (perhaps related to neuronal populations or pathways with particular selectivity, e.g. to certain facial features), different 'highways' might dominate in different conditions and be more or less amenable to control by task-related factors. Therefore, some evaluative processing in the amygdala may be unconscious, uncontrolled, independent of voluntary attention, efficient, and/or fast (Moors and De Houwer, 2006), and hence take place without any explicit task requirements, which could then serve to facilitate attention, learning and memory. However, for other evaluative processes and/or other conditions, current task demands might still exert an important modulation on amygdala responses, even when emotion is irrelevant. Future research should further clarify whether responses to faces presented in unconscious or implicit conditions may reflect their (un)trustworthy traits as judged consciously during a more explicit task.

Finally, what did we learn on trust? Although this attribute is among the most common judgments spontaneously made on unknown faces (Oosterhof and Todorov, 2008), the distinctive properties implied by such judgments are not straightforward to define. Though usually associated with components of social and economic exchanges (Fehr and Zehnder, in press), trust essentially denotes a situation of perceived security and thus directly contrasts with threat, i.e. the most relevant emotional signal typically responsible for eliciting amygdala activation. Thus, to some extent, it may come as no surprise that (un)trustworthiness constitutes a key attribute of faces to which the amygdala appears particularly sensitive. However, the exact facial information underlying consensual judgments of trustworthiness on isolated, static pictures of unknown faces has remained rather elusive. The recent work of Todorov and colleagues has also begun to shed some valuable light on this issue, by showing that the prototypical features of untrustworthy faces (as judged in static pictures without context) share several attributes with angry expression (Oosterhof and Todorov, 2008), which may then also explain a preferential amygdala activation to these faces. However, in real life, trust is likely to involve several other dimensions that may strongly depend on factors other than just the particular physical features of faces (Buckingham *et al.*, 2006), including not only context but also past experiences, cultural stereotypes including gender and racial traits, non-facial cues such as hairstyle or clothes, as well as current goals for social relationships. For example, even though many people would claim not to trust a face with frowning eyes and a prognathous unshaven chin when they lack any other information (as the students tested by Oosterhof and Todorov, 2008), it is likely that the same people would prefer to trust this face more than a slim warmly smiling woman if asked to choose a bodyguard to travel to some dodgy neighborhood in Los Angeles. Likewise, the perceived trustworthiness of a face may differ for a surgeon or a secretary (Glanville and Paxton 2008; Gorn *et al.*, 2008). Therefore, there might be some limitations to the study of complex social judgments such as trustworthiness out of a richer situational context. Future studies might usefully manipulate social goals during trustworthiness (or other trait) judgments, in order to go beyond the identification of general features that are highly correlated with simple facial features such as the components of basic emotion expressions (e.g. anger- or disgust-like facial action unit patterns, see Oosterhof and Todorov, 2008).

In sum, a reason why studies of trust have proven so fruitful for dissecting the mechanisms of face perception and emotional appraisal might be that this crucial and finely tuned component of our spontaneous decisions about others may well capture several of the most essential ingredients of the human amygdala contribution to social processing: this may include not only the detection of potential threat, but also the evaluation of self-relevance and of a more general valence or value of ongoing events, as well as their uncertainty and importance for subsequent learning.

ACKNOWLEDGEMENTS

D.S. and P.V. are supported by the Swiss National Science Foundation (FNRS grant No 105314-118267) and by the Swiss National Centre of Competence in Research (NCCR) in Affective Sciences (grant No 51NF40-104897).

Patrik Vuilleumier[1,2,3] and David Sander[1,2,4]
[1]Neuroscience Center,
[2]Swiss Center for Affective Sciences,
[3]School of Medicine, and
[4]Department of Psychology, University of Geneva, 1205 Geneva, Switzerland

REFERENCES

Adolphs, R., Tranel, D. (2004). Impaired judgments of sadness but not happiness following bilateral amygdala damage. *Journal of Cognitive Neuroscience*, *16*, 453–62.

Amaral, D. (2003). The amygdala, social behavior, and danger detection. *Annals of the New York Academy of Sciences*, *1000*, 337–47.

Anderson, A.K., Christoff, K., Stappen, I., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience*, *6*, 196–202.

Blair R.J., Morris J.S., Frith C.D., Perrett D.I., Dolan R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, *122*(Pt 5), 883–93.

Brosch, T., Moors, A. (in press). Valence. In: Sander, D., Scherer, K.R., editors. *The Oxford Companion to Emotion and the Affective Sciences*. Oxford and New York: Oxford University Press.

Buckingham, G., DeBruine, L.M., Little, A.C., et al. (2006). Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. *Evolution and Human Behavior*, *27*, 381–9.

Cacioppo, J.T., Larsen, J.T., Smith, N.K., Berntson, G.G. (2004). The affect system: What lurks below the surface of feelings. In: Manstead, A.S.R., Frijda, N., Fischer, A., editors. *Feelings and Emotions*. Cambridge, UK: Cambridge University Press, pp. 221–40.

Colombetti G. (2005). Appraising Valence. *Journal of Consciousness Studies*, *12*(8–10).

Cunningham, W. A., Van Bavel, J. J., Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psychological Science*, *19*, 152–60.

Fehr, E., Zehnder, C. (in press). Trust. In: Sander, D., Scherer, K.R., editors. *The Oxford Companion to Emotion and the Affective Sciences*. Oxford and New York: Oxford Universty Press.

Fitzgerald, D.A., Angstadt, M., Jelsone, L.M., Nathan, P.J., Phan, K.L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *Neuroimage*, *30*, 1441–8.

Glanville, JL, Paxton, P. (2008). How do we learn to trust? A confirmatory tetrad analysis of the sources of generalized trust. *Social Psychology Quarterly*, *70*, 230–42.

Gorn, G.J., Jiang, Y., Johar, G.V. (2008). Babyfaces, trait inferences, and company evaluations in a public relations crisis. *Journal of Consumer Research*, *35*, 36–49.

Hamann, S. (2003). Nosing in on the emotional brain. *Nature Neuroscience*, *6*, 106–8.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., et al. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience*, *22*, 2730–6.

Moors, A., De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297–326.

Öhman, A., Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*, 483–522.

Oosterhof, N. N., Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA*, *105*, 11087–92.

Ousdal, O.T., Jensen, J., Server, A., et al. (2008). The human amygdala is involved in general behavioral relevance detection: Evidence from an event-related functional magnetic resonance imaging Go-NoGo task. *Neuroscience*, *156*, 450–5.

Phelps, E.A., LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, *48*(2), 175–87.

Pourtois, G., Schwartz, S., Spiridon, M., Martuzzi, R., and Vuilleumier, P. (in press) Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cerebral Cortex*.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*, 148–158.

Russell, J.A., Carroll, J.B. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3–30.

Sander, D., Grafman, J., Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*, 303–16.

Sander, D., Grandjean, D., Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*, 317–52.

Schaefer, A., Braver, T. S., Reynolds, J. R., Burgess, G. C., Yarkoni, T., Gray, J. R. (2006). Individual differences in amygdala activity predict response speed during working memory. *The Journal of Neuroscience*, *26*, 10120–8.

Sergerie, K., Chochol, C., Armony, J.L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *32*, 811–30.

Vuilleumier, P. (in press). The role of the amygdala in perception and attention. In Whalen, P.J., Phelps, E.A., editors. *The Human Amygdala*. New York: Guilford Press.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Science*, *9*, 585–94.

Vuilleumier, P., Armony, J., Dolan, R. (2003). Reciprocal links between emotion and attention. In: Frackowiak, R.S.J., et al., editors. *Human Brain Function*. San Diego: Academic Press, pp. 419–44.

Whalen, P.J., Kagan, J., Cook, R.G., et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science*, *306*, 2061.

Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., Wright, C.I., Rauch, S.L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion*, *1*, 70–83.

Winston, J.S., Gottfried, J.A., Kilner, J.M., Dolan, R.J. (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *Journal of Neuroscience*, *25*, 8903–7.

Winston, J.S., Strange, B., O'Doherty, J., Dolan, R. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of face. *Nature Neuroscience*, *5*, 277–83.

Wright, P., Liu, Y. (2006). Neutral faces activate the amygdala during identity matching. *Neuroimage*, *29*, 628–36.

Winston, J.S., Gottfried, J.A., Kilner, J.M., Dolan, R.J. (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *J Neurosci*, *25*(39), 8903–7.