

How plausible is a subcortical account of rapid visual recognition?

Maxime Cauchoix^{1,2*} and Sébastien M. Crouzet^{3,4}

¹ Centre de Recherche Cerveau et Cognition, Université Paul Sabatier, Université de Toulouse, Toulouse, France

² Faculté de Médecine de Purpan, CNRS, UMR 5549, Toulouse, France

³ Department of Cognitive, Linguistic and Psychological Science, Brown University, Providence, RI, USA

⁴ Institute of Medical Psychology, Charité University Medicine, Berlin, Germany

*Correspondence: cauchoix@cerco.ups-tlse.fr

Edited by:

Josef Parvizi, Stanford University, USA

Reviewed by:

Josef Parvizi, Stanford University, USA

Ueli Rutishauser, California Institute of Technology, USA

Ido Davidesco, Hebrew University of Jerusalem, Israel

Primates recognize objects in natural visual scenes with great rapidity. The ventral visual cortex is usually assumed to play a major role in this ability (“high-road”). However, the “low-road” alternative frequently proposed is that the visual cortex is bypassed by a rapid subcortical route to the amygdala, especially in the case of biologically relevant and emotional stimuli. This paper highlights the lack of evidence from psychophysics and computational models to support this “low-road” alternative. Most importantly, the timing of neural responses invites a serious reconsideration of the low-road role in rapid processing of visual objects.

THE SPEED OF SIGHT

The rapid and accurate processing of complex visual scenes has been demonstrated by Thorpe and colleagues using the rapid visual categorization protocol (Thorpe et al., 1996), in which participants reported the presence of animals in natural scenes as soon as 250 ms after image onset. This result sets strong time constraints on the neural mechanisms underlying object categorization. Diagnostic category information might actually be available even earlier, since selective eye movement responses can be produced only 100–120 ms after stimulus onset (Kirchner and Thorpe, 2006; Crouzet et al., 2010). What neural mechanisms could account for such rapid vision?

THE CORTICAL “HIGH-ROAD”

A widely held view is that object recognition results from the interplay

of hierarchically organized areas along the ventral visual stream (Dicarlo et al., 2012)—running from the primary visual cortex (V1) through extrastriate visual areas (V2 and V4), to the inferotemporal cortex (IT) where high-level visual representations are encoded. To reconcile this view with the short behavioral latencies observed in rapid categorization tasks, several authors have suggested that a pure feedforward sweep of activity through the ventral stream might be sufficient to perform core object recognition (Thorpe et al., 1996; Serre et al., 2007b).

THE SUBCORTICAL “LOW-ROAD”

On the other hand, a subcortical shortcut—the so-called “low-road”—might seem to be a plausible alternative. This hypothesis finds its origin in the rapid amygdala responses reported by (LeDoux, 1996) during auditory fear conditioning. In a series of experiments in rodents, he delineated a quick route that bypasses the cortex by directly reaching the amygdala via the thalamus. Such a subcortical shortcut would, in specific cases such as threatening situations, enable the rapid initiation of appropriate defense responses even before the sensory cortices become involved. Furthermore, since the amygdala has been linked to emotion recognition (particularly fear) in humans (Adolphs et al., 1994), this alternative pathway was proposed as an explanation for rapid, automatic, and unconscious reactions among humans and monkeys to biologically relevant visual stimuli (Öhman and

Mineka, 2001; Johnson, 2005; Öhman, 2005; Vuilleumier, 2005; Tamietto et al., 2009; Tamietto and de Gelder, 2010; de Gelder et al., 2011).

Here we argue that there is no convincing evidence in support of the “low-road” theory when extended to rapid visual object processing in primates. To preface our arguments, first, real-world object categorization requires computational properties that have not yet been found in a subcortical pathway (see “Real-world Recognition Requires Selectivity and Invariance”). Second, among the characteristics attributed to the “low-road,” we argue that genuine rapidity has not yet been demonstrated appropriately (see “What is Rapid Visual Processing?”). Finally, we will demonstrate how the low-road hypothesis is at odds with neural latencies reported in the amygdala and the visual cortex (see “Ventral Stream Visual Cortex is Activated Before the Amygdala”). Altogether, these arguments point to an earlier role for cortical areas and suggest a serious reconsideration of the role of the “low-road” in rapid vision.

REAL-WORLD RECOGNITION REQUIRES SELECTIVITY AND INVARIANCE

To support recognition, a neural system needs to reach a high level of selectivity while dealing with the inherent variability of sensory input. This balance between selectivity and invariance is a hallmark feature of visual recognition in primates, and remains a challenge for computer vision.

In macaque monkeys, selective neural responses to complex objects are typically found in the IT (Dicarlo et al., 2012). These neuronal responses are also tolerant to changes in retinal position, scale, or pose of the object (Hung et al., 2005). Studies using intracranial recordings in human epileptic patients have also shown that neural responses from the visual cortex provide a categorical signal tolerant to changes in scale and position (Liu et al., 2009).

Driven by results from electrophysiology, a plausible model of how selectivity and invariance could be built through the ventral stream has emerged. It is based on two successive operations, template-matching and non-linear pooling, repeated at each stage of the ventral hierarchy (Serre et al., 2007b). Such hierarchical models have been shown to accurately mimic primate rapid categorization performance (Serre et al., 2007b; Crouzet and Serre, 2011) and neural responses of the visual ventral stream (Serre et al., 2007a).

Among subcortical structures, human single-unit studies showed that the amygdala contains neurons selective to categories or objects such as animals, famous faces, or places (Kreiman et al., 2000; Quiroga et al., 2005; Mormann et al., 2011). Interestingly, these neurons are highly invariant since they respond to various pictures of their preferential objects, but also to their written or spoken names. However, there is currently no model of how this high level of both selectivity and invariance could be built from a direct subcortical route. A more reasonable assumption would thus be that it gets its input from high-level areas of the ventral stream, rather than from the thalamus (shortcut “low-road” route).

WHAT IS RAPID VISUAL PROCESSING?

Numerous studies investigated affective stimulus processing with short image presentation and masking protocols to show that emotions such as fear can be processed unconsciously and “rapidly” (Bar et al., 2006; Öhman et al., 2007; Adolphs, 2008). While there is no doubt that masking is a powerful experimental tool to reveal unconscious sensory processing, it does not provide information on the genuine rapidity of visual processing. In

backward masking protocols, the stimulus onset asynchrony (SOA, time interval between target and mask onset) is a measure of the visual uptake time (or temporal resolution), not of the time required for complete visual processing. In other words, even in a perfect pipeline model of the visual system, the mask interference would only give information about the time spent at each stage, and not about the cumulative time for all stages (Vanrullen, 2011). For example, the fact that fear information can be extracted from faces masked after an SOA of 39 ms (Bar et al., 2006) informs us about the minimal visual uptake time necessary for fear processing but does not say anything about the time at which fear information is available to trigger behavioral responses.

The speed of processing for object or scene categorization has been extensively studied using rapid categorization protocols. Using minimal reaction time measurements (the time at which correct responses start to significantly outnumber incorrect ones) it has been shown that humans can categorize images as containing an animal in only 250 ms (Thorpe et al., 1996), while monkeys can perform the same task by 180 ms (Fabre-Thorpe et al., 1998). Even faster, reliable saccades toward faces and animals can be triggered as soon as 100–120 ms after image onset (Kirchner and Thorpe, 2006; Crouzet et al., 2010). As far as we know, there is no evidence for faster processing of emotional stimuli, as would be predicted by the “low-road” hypothesis.

VENTRAL STREAM VISUAL CORTEX IS ACTIVATED BEFORE THE AMYGDALA

Most of the studies on humans investigating the role of the amygdala in visual processing used fMRI and PET scans (Morris et al., 1999; Whalen et al., 2004; see Pessoa and Adolphs, 2010 and Vuilleumier, 2005 for reviews). These two techniques, because of their poor temporal resolution, do not allow conclusions about the temporal dynamics of stimulus processing. Despite this limitation, it was assumed that amygdala responses to emotional stimuli, notably to fear-inducing stimuli, were based on a rapid low-road activation (Öhman and Mineka, 2001; Vuilleumier, 2005).

A review of electrophysiological studies reporting neural latencies suggests a clearly different picture. Many studies investigating the properties of IT cells have reported selective responses to shapes, faces or object categories occurring as soon as 70–100 ms after stimulus onset (Perrett et al., 1982; Li et al., 1993; Tovee et al., 1994; Liu and Richmond, 2000; Hung et al., 2005; see Mormann et al., 2008 for a review). Similarly, in human epileptic patients, IFP recorded from the occipito-temporal cortex were object category selective as early as 100 ms after stimulus onset (Liu et al., 2009). These category selective latencies are compatible with the rapid behavioral responses observed in natural scene categorization tasks (Thorpe et al., 1996; Fabre-Thorpe et al., 1998; Kirchner and Thorpe, 2006; Girard et al., 2008). Furthermore, this early ventral stream activity has been shown to be causally linked with behavioral responses in monkeys (Afraz et al., 2006) and humans (Pitcher et al., 2007; Sadeh et al., 2011).

On the other hand, selective responses to visual features or objects in monkeys’ amygdala tend to have a greater time-lapse (Gothard et al., 2007). One single-unit study (Leonard et al., 1985) compared the two pathways directly (on the same monkeys) and showed that neurons in the STS (superior temporal sulcus, top of the ventral stream) had latencies (90–140 ms) that systematically preceded those from the amygdala (110–200 ms). In humans, two intracranial recording studies have tested the existence of rapid amygdala responses to emotional stimuli. But the earliest responses were reported at 200 ms (Krolak-Salmon et al., 2004) and 250–500 ms (Rutishauser et al., 2011), which is much slower than the fast occipito-temporal selectivity reported for objects categories (Liu et al., 2009). Moreover, the amygdala responses observed for emotional stimuli were not occurring earlier than what is generally reported for object categories (Mormann et al., 2008). Among the medial temporal lobe structures (i.e., perirhinal cortex, entorhinal cortex, hippocampus, and amygdala), the amygdala is actually the one with the slowest visual responses (average latencies of 271 ms in the perirhinal cortex for example). The pattern of

neural latencies observed in both human and monkeys thus clearly vouches for a cortical “high-road” precedence.

CONCLUSION

In this paper we questioned the hypothesis that a subcortical low-road could account for the speed of sight. Several observations from psychophysics, computational modeling, and electrophysiology strongly suggest that the low-road account is mostly incompatible with the characteristics of rapid visual categorization. On the contrary, a large collection of evidence confirms that the cortical high-road, through the visual ventral stream, can accomplish a rapid, selective, and invariant analysis of the scene. The latency of neural visual activation and response characteristics in the amygdala clearly suggest that its involvement in visual processing is downstream of the ventral visual cortex, after core object recognition has been performed.

Thus, contrary to what is commonly acknowledged, rapid, and automatic processing of visual objects is likely to be under cortical-dependence while subcortical structures would be involved in slower (probably higher-level) processing. This conclusion conforms with recent results and reviews pointing out the unexpected role of subcortical structure in high cognitive functions (Parvizi, 2009). Amygdala for instance is now thought to play a major role in the evaluation of the biological significance of stimuli (Pessoa and Adolphs, 2010) and the pulvinar, showing dense connection with many cortical areas, has recently been shown to play a role in regulating information transmission across the visual cortex (Saalmann et al., 2012).

ACKNOWLEDGMENTS

We would like to thanks Ralph Adolphs, Ali Arslan, Ramakrishna Chakravarthi, Julien Dubois, Katalin Gothard, Gabriel Kreiman, Marianne Latinus, Edmund Rolls, Imri Sofer, Leslie Ungerleider, and Rufin VanRullen for their encouragement, comments, and suggestions on this manuscript.

REFERENCES

- Adolphs, R. (2008). Fear, faces, and the human amygdala. *Curr. Opin. Neurobiol.* 18, 166–172.
- Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372, 669–672.
- Afraz, S.-R. S., Kiani, R. R., and Esteky, H. H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature* 442, 692–695.
- Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6, 269–278.
- Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *J. Vis.* 10, 16.1–16.17.
- Crouzet, S. M., and Serre, T. (2011). What are the visual features underlying rapid object recognition? *Front. Psychology* 2:326. doi: 10.3389/fpsyg.2011.00326
- de Gelder, B., van Honk, J., and Tamiotto, M. (2011). Emotion in the brain: of low roads, high roads and roads less travelled. *Nat. Rev. Neurosci.* 12, 425. author reply: 425.
- Dicarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434.
- Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9, 303–308.
- Girard, P., Jouffrais, C., and Kirchner, C. H. (2008). Ultra-rapid categorisation in non-human primates. *Anim. Cogn.* 11, 485–493.
- Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitzer, K. M., and Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* 97, 1671–1683.
- Hung, C., Kreiman, G., Poggio, T. A., and Dicarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866.
- Johnson, M. H. (2005). Subcortical face processing. *Nat. Rev. Neurosci.* 6, 766–774.
- Kirchner, H., and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res.* 46, 1762–1776.
- Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci.* 3, 946–953.
- Krolak-Salmon, P., Hénaff, M.-A., Vighetto, A., Bertrand, O., and Mauguière, F. (2004). Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: a depth electrode ERP study in human. *Neuron* 42, 665–676.
- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, 1st Edn. New York, NY: Simon and Schuster.
- Leonard, C. M., Rolls, E. T., Wilson, F. A., and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behav. Brain Res.* 15, 159–176.
- Li, L., Miller, E. K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* 69, 1918–1929.
- Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290.
- Liu, Z. Z., and Richmond, B. J. B. (2000). Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *J. Neurophysiol.* 83, 1677–1692.
- Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., et al. (2011). A category-specific response to animals in the right human amygdala. *Nat. Neurosci.* 14, 1247–1249.
- Mormann, F., Kornblith, S., and Quiroga, R. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J. Neurosci.* 28, 8865–8872.
- Morris, J. S., Öhman, A., and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating “unseen” fear. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1680–1685.
- Öhman, A. (2005). The role of the amygdala in human fear: automatic detection of threat. *Psychoneuroendocrinology* 30, 953–958.
- Öhman, A., Carlsson, K., Lundqvist, D., and Ingvar, M. (2007). On the unconscious subcortical origin of human fear. *Physiol. Behav.* 92, 180–185.
- Öhman, A., and Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychol. Rev.* 108, 483–522.
- Parvizi, J. (2009). Corticocentric myopia: old bias in new cognitive sciences. *Trends Cogn. Sci.* 13, 354–359.
- Perrett, D. I., Rolls, E. T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* 47, 329–342.
- Pessoa, L., and Adolphs, R. (2010). Emotion processing and the amygdala: from a “low road” to “many roads” of evaluating biological significance. *Nat. Rev. Neurosci.* 11, 773–783.
- Pitcher, D., Walsh, V., and Yovel, G. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Curr. Biol.* 17, 1568–1573.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107.
- Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A. N., Heller, A. C., Ross, I. B., et al. (2011). Single-unit responses selective for whole faces in the human amygdala. *Curr. Biol.* 21, 1654–1660.
- Saalmann, Y. B. Y., Pinsk, M. A. M., Wang, L. L., Li, X. X., and Kastner, S. S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756.
- Sadeh, B., Pitcher, D., Brandman, T., Eisen, A., Thaler, A., and Yovel, G. (2011). Stimulation of category-selective brain areas modulates ERP to their preferred categories. *Curr. Biol.* 21, 1894–1899.
- Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., and Poggio, T. A. (2007a). A quantitative theory of immediate visual recognition. *Prog. Brain Res.* 165, 33.
- Serre, T., Oliva, A., and Poggio, T. A. (2007b). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429.

- Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., et al. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17661–17666.
- Tamietto, M., and de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nat. Rev. Neurosci.* 11, 697–709.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.
- Tovee, M. J. M., Rolls, E. T. E., and Azzopardi, P. P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *J. Neurophysiol.* 72, 1049–1060.
- Vanrullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Front. Psychology* 2:365. doi: 10.3389/fpsyg.2011.00365
- Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci.* 9, 585–594.
- Whalen, P., Kagan, J., Cook, R., and Davis, F. (2004). Human Amygdala Responsivity to masked fearful eye whites. *Science* 306, 2061.
- Received: 17 November 2012; accepted: 03 February 2013; published online: 27 February 2013.
- Citation: Cauchoix M and Crouzet SM (2013) How plausible is a subcortical account of rapid visual recognition? *Front. Hum. Neurosci.* 7:39. doi: 10.3389/fnhum.2013.00039
- Copyright © 2013 Cauchoix and Crouzet. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.