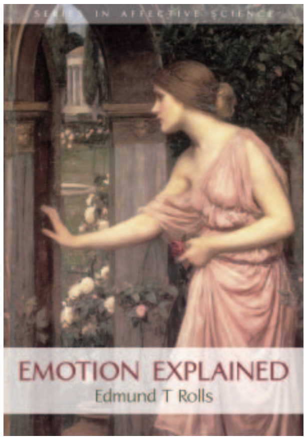
Brain (2006), **129**, 2517–2521

BOOK REVIEW

Explaining emotion in the brain

EMOTION EXPLAINED
By *Edmund T. Rolls*
2005. Oxford: Oxford
University Press
Price: £39.95
ISBN: 0-19-857003-1



Emotion has become a well-represented, and well-respected, topic of neuroscientific inquiry, as borne out by its exponential increase in citation indices (e.g. increases in publications with the word in their title by 300 in the 1980s, by 900 in the 1990s and by >1600 in the past 6 years according to Science Citation Index). The reasons for this increase lie with advances both in our theoretical understanding of emotions and in the development of new tools for exploring their neural basis—notably functional neuroimaging. These efforts have resulted in important advances also in the diagnosis and treatment of neurological and psychiatric diseases, many of which feature dysfunctional emotions as one of the most disabling components. It is thus timely to take stock of what we have learned, and to publish books that summarize the state of the field.

Edmund Rolls' latest book, *Emotion Explained*, provides such a summary. The book opens with a series of questions: 'What are emotions? Why do we have emotions?. . . Why does it feel like something to have an emotion?' How good an answer the next 450 pages and two appendices provide will depend on the kinds of explanations the reader finds most satisfying. If you like explanations that aim towards a vocabulary of genes, neurons and reinforcers, then the book is for you. But if you find macroscopic or socially contextualized accounts more appealing, you should still read the book, both for its wealth of facts from behavioural neuroscience and for its clear statement of positions with which you will disagree.

Rolls is Professor of Experimental Psychology at Oxford, and his laboratory has provided one of the most sustained and comprehensive neurophysiological investigations of associative memory, higher-level sensory processing and motivation of any group in the world. The studies provide much of the empirical foundation for understanding emotions, especially in regard to the orbitofrontal cortex and amygdala, two structures whose functions were reviewed in extensive detail in Rolls' earlier book, *The Brain and Emotion* (Rolls, 1999). The present sequel provides 11 chapters of extensions and updates (the differences from the previous book are summarized in a preface) to a theoretical and empirical review of Rolls' work on emotion. As with the earlier book, the answer to the question, 'What are emotions?' is an expanded account of how emotions are caused by reward or punishment. Hunger, thirst and sex, each occupying a whole chapter, provide the foundational examples on which more complex emotions can be built. The emphasis is on reinforcement learning: how associations are acquired and stored in the brain between representations of sensory stimuli and representations of their reinforcement value. A nice 46-page appendix covers computational models of associative memory for further background.

The book is chock-full of interesting and updated facts, and provides discussions of hunger (the longest section), thirst, brain-stimulation reward, addiction and sexual behaviour (a compendium of entertaining factoids explained in the style of evolutionary psychology) along the way. Somewhat surprisingly, given the computational modelling already advocated in the book, only a very short section at the end provides a link with behavioural economics, a discipline whose mathematical analysis of decision-making is becoming an increasingly popular adjunct to neuroscience investigations and has spawned the new discipline of 'neuroeconomics' (Glimcher, 2003). It would have been interesting to know more about how Rolls sees the relation between models of reinforcement learning and of microeconomic choice.

Another prominent opening question about emotions is 'What is their adaptive value?' Here Rolls takes the somewhat extreme view that all emotions have arisen for evolutionary reasons, and that the behavioural goals on which they are based are entirely specified in the genome. Echoing Dawkins (whose book, *The Selfish Gene*, is cited in both its first and second editions), selfish genes are seen as the unit of natural selection, and as containing all of the information about the goals whose achievement would lead to their reproduction. I like the idea that emotions encapsulate a motivation to achieve particular, adaptive goals, rather than the specific behaviours, a view broadly

© The Author (2006). Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

consonant with appraisal theories. But I don't understand why all of this should be specified in the genome, nor how claiming so really explains much in the context of the book. The claim would seem to exclude goals that are specified, entirely through a different mechanism for heritability, cultural learning. (There is a puzzling and extremely brief remark on p. 62 that seems to acknowledge that not all goals are genetically specified, but this seems at odds with statements elsewhere in the book.) While genes may be the ultimate target of natural selection, I also am not sure that they are the basic units of such selection, which might well be individuals or indeed groups in highly social species. Well-known puzzles about non-reciprocal altruism in some animals (notably humans) may best be explained by assuming that the unit of selection is the group, not the gene (contrary to the blurb on social attachment given on p. 57 of the book). Views on these topics are highly polarized and the literature often technical; at this stage their link to emotion and the need for any such link in the book seem unclear.

Most of the data in the book come from single-unit neurophysiological studies, many of them conducted by Rolls and his students, although there is also a substantial presentation of neuroimaging data and some lesion studies. The book correctly points out the coarse resolution of functional neuroimaging methods, and I agree that 'the presence of [fMRI] blobs should not be taken as more than a gross reflection of the underlying neuronal representations. . . .' (p. 128). But writing that 'such imaging techniques give rather little evidence on how the brain works' (p. 6) is an overstatement that is somewhat puzzling in light of the substantial amount of imaging data that is in fact reviewed in the book—much of it from Rolls' own lab (some 13 plates are provided in the colour insert of the book). Just as the unit of natural selection is taken to be the gene, the level of explanation of brain function is taken to be the neuron. But this is not so clear. The level of explanation may be neuronal, but it would have to be found in the distributed pattern of many neurons whose activity could be read out by other neurons to which they project—a daunting task in a brain of any complexity (it has been achieved very elegantly in some much simpler invertebrate sensory systems). Tables 4.2 and 4.3 in the book detail over a dozen different ways of classifying the response properties of neurons in the orbitofrontal cortex. Which of these, or which combinations of them, should be designated as the brain's 'basis functions' by which emotions are represented? The experimentalist's intuitive classification scheme from looking at the discharge rates of the neurons may well be irrelevant to this task. Rolls is aware of these difficulties and stresses the distributed, population-level nature of neuronal coding.

The approach of Rolls' theory of emotions can be gleaned from Figure 2.1 and Table 2.1 in the book (*see* Fig. 2.1). The withholding or administration of positive or negative reinforcement defines emotional states. Rolls takes care to point out that this is not the same as dimensional psychological theories of emotion, but rather intended to provide a framework for locating emotions, at varying levels of intensity, in a space of four possibly independent dimensions. The view is also much more sophisticated in many respects than classical Behaviourism, and broadly consists of three stages in which an organism acquires knowledge of the value of stimuli in the world. The first stage is phylogenetically specified: genes set the goals (reward and punishment) towards which organisms must direct their actions, as already noted above. The second stage is a mixture of innate and acquired stimulus–reward associations—learning the value of primary and secondary reinforcers, phylogenetically for the former, and through stimulus–stimulus associations for the latter. And the third stage involves mostly experience-dependent instrumental learning of how to achieve the goals specified in the first stage through one's actions (action–outcome learning). The flexible nature of such instrumental learning is emphasized: emotions specify the goals to be obtained, but not the actions to obtain them; likewise for the genes. Rolls' view also departs notably from classical Behaviourism in that it discusses how top-down cognitive factors, and indeed consciousness of one's plans and goals, can influence emotion and behaviour. This is treated in detail in later chapters of the book, which develop a 'dual routes to action' account that acknowledges an explicit route that involves planning and deferred reward.

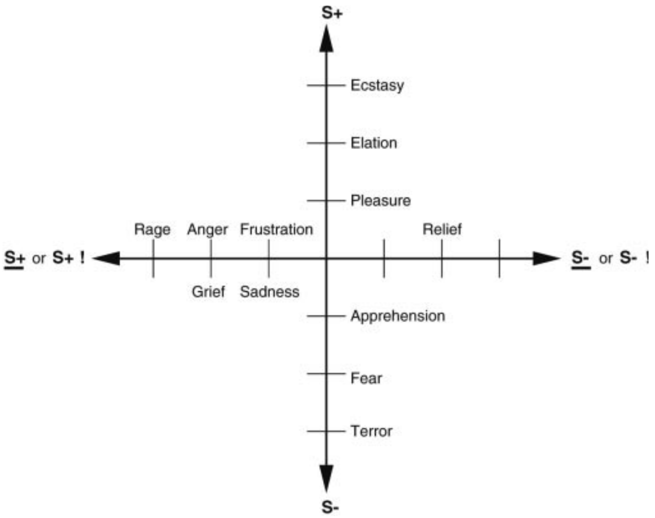


Fig. 2.1 Framework for understanding specific emotions as arising from specific reinforcement contingencies. The scheme is not meant to be a dimensional model for emotions, and the four axes could in principle be independent. From "Emotion Explained" by Rolls, Edmund (2005) reproduced with permission.

The book contains a separate chapter on a theory of consciousness, where explicit planning and the role of feelings are discussed more. As Rolls himself acknowledges, the theory of consciousness is speculative, as all theories of consciousness must be. It is a 'higher-order thought' theory, similar to what has been proposed by philosophers such as David Rosenthal. However, Rolls' theory of consciousness is distinguished by an emphasis on symbolic meta-representation. While this need not be in terms of language as such, it is like language in that it allows an organism to think about its behaviour in a way more abstract than what implicit cognition could provide. It is such symbolic representation (typically in terms of language in humans) that gives an organism conscious awareness of its emotions.

There is much to be said for the view on emotion offered in this book: it is probably the most empirically tractable and the most mathematically precise framework for understanding emotion we currently have available, borne out by the vast and growing numbers of studies on the topic. It is consistent with the phylogenetic expansion of forebrain tissue required for storing the associations to guide the more flexible behaviours of mammals—notably the invention of neocortex. It is based on solid electrophysiological studies of basic motivational processing. The book has sections that articulate the virtues of these ideas and how they can explain a broad range of data (there is a subsection entitled, 'Advantages of the approach to emotion described here'). But how does it relate to other theories of emotion?

The book mentions several other emotion theories in passing (in six pages), but these are treated superficially—perhaps understandably so, given the volume of material already devoted to advancing the author's own and to reviewing the neuroscience data. Nonetheless, it would have been nice to see some additions and amendments. For example, the writings of the psychologist Klaus Scherer are mentioned briefly, but no reference is made to his theory according to which different components of stimulus evaluation are generated by the brain at different points in time (Scherer, 1999). Indeed, while the spatially distributed nature of neuronal representations is well treated in the book, their temporal dispersion is hardly covered and the dynamical nature of emotion somewhat neglected. Relatedly, there is no discussion of event-related brain potential (ERP) studies of emotion.

Appraisal theorists like Scherer, Lazarus and others are often difficult for neuroscientists to understand because of their emphasis on relations between an organism's goals, plans and coping strategies rather than on simpler stimulus–response associations. However, some writers, such as the philosopher Jesse Prinz, have incorporated appraisal theory with evolutionary and neural accounts to give what amounts to a translation between the different frameworks for thinking about emotion (Prinz, 2004). A more elaborated discussion of other emotion theories in *Emotion Explained* could have provided a basis for drawing in an even broader readership to whom this book will be relevant.

A continuing frustration to me, in this book as in his other writings, is with Rolls' misunderstanding of Antonio Damasio's somatic-based theory of emotion (Damasio, 1994), updated here with a discussion of the equally misguided debate between Damasio and Jay McClelland concerning the 'somatic marker hypothesis' (Damasio's theory of decision-making, a related but distinct topic). The details are covered elsewhere [pp. 26–30 of the book, and (Bechara *et al.*, 2005)] and not so important here, but there is a foundational issue that is very important and my only real criticism of Rolls' theory of emotion. Damasio emphasizes that representations of an organism's homeostatic state (or deviations thereof) ground emotions. (In his theory, they ground a particular component of emotions he calls 'feelings', although these need not be conscious.) Rolls disagrees with this view and thinks that what grounds emotions are the value-representations stored in structures like the orbitofrontal cortex. This is puzzling, because Rolls devotes some time (pp. 66–90) to discussing the neural basis for representing the sensory properties of primary and secondary reinforcers, and correctly emphasizes that it would be inefficient for the brain simply to associate representations at the level of the retina, since the retina does not yet represent information at the right level of objects that can be generalized over different viewpoints, locations and exemplars. Now, why not follow the same design principles at the other end of the stimulus–value association? If value is a homeostatic goal state of the organism, as Rolls acknowledges, then we should find an equally hierarchical processing stream coming into the orbitofrontal cortex from interoceptive representations of the organism's homeostatic state. If neurons in the orbitofrontal cortex associate two things—the sensory properties of stimuli and their reinforcing properties—then we need to have representations of those two things in the brain such that they can be brought together. The description of the sensory end, as noted, is adequately sophisticated. But the reinforcing end is rather left dangling, apparently under the assumption that everything is there in the orbitofrontal cortex. But how could it be, without a chain of inputs that grounds it in what the reinforcement is all about: the homeostatic state of the organism. Rolls sees the body itself merely as one of the possible, and in the end essentially epiphenomenal, routes of output for emotion (e.g. autonomic and endocrine responses are a 1-page subsection under 'Further functions of emotion', p. 51). Damasio sees it as the grounding input for emotion.

Figures 4.2 and 4.3 in Rolls' book summarize how secondary reinforcers can be associated with primary reinforcers, but there is no account of how primary reinforcers can be represented in the first place (*see* Fig. 4.3). There is the sensory input about the stimulus, and there are orbitofrontal cortex and amygdala that store the association between stimulus and value, but there is no place where value is represented and can be grounded. Such a place could be found in the insula and the brainstem nuclei, which writers such as Damasio, Jaak Panksepp and Bud Craig discuss at some length. The omission to develop this concept in *Emotion Explained* is all the more puzzling given the very extensive chapter on hunger, which details phenomena such as sensory-specific satiety: Rolls' seminal

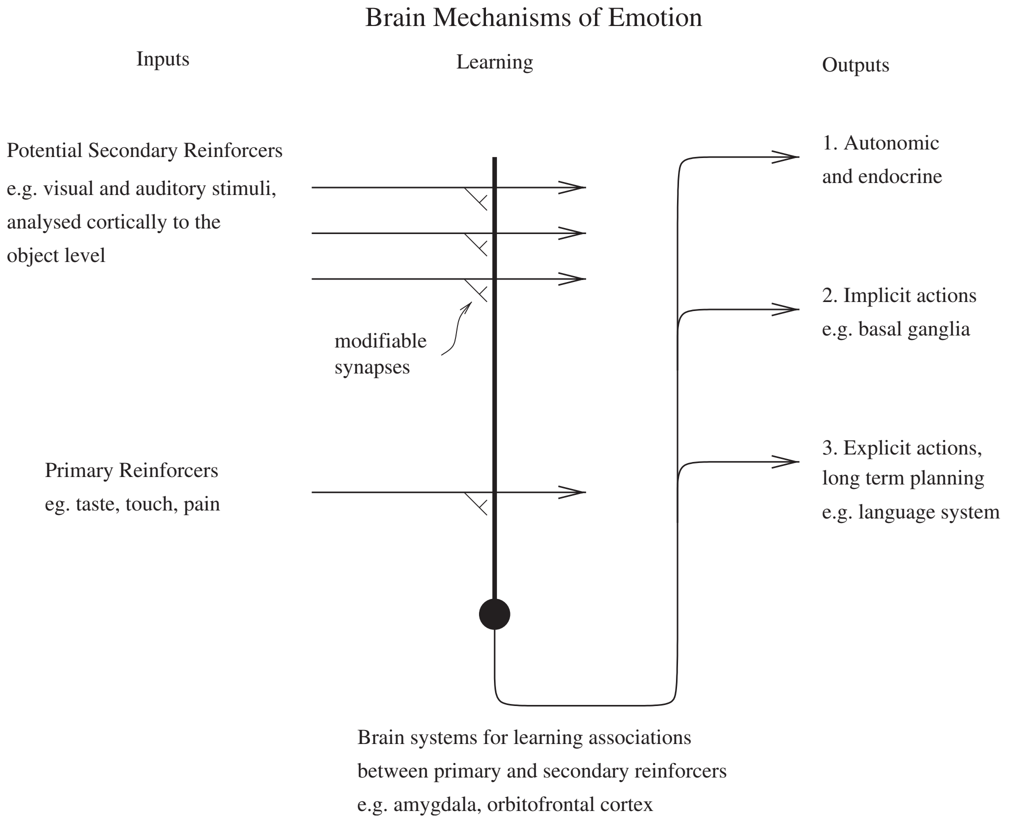

Brain Mechanisms of Emotion
Inputs
Learning
Outputs
Potential Secondary Reinforcers
e.g. visual and auditory stimuli, analysed cortically to the object level
modifiable synapses
Primary Reinforcers
eg. taste, touch, pain
1. Autonomic and endocrine
2. Implicit actions
e.g. basal ganglia
3. Explicit actions, long term planning
e.g. language system
Brain systems for learning associations between primary and secondary reinforcers
e.g. amygdala, orbitofrontal cortex

Fig. 4.3 Schematic diagram showing the organization of some of the brain mechanisms underlying emotion, including those involved in learning the reinforcement associations of visual stimuli. From "Emotion Explained" by Rolls, Edmund (2005) reproduced with permission.

discovery that neurons in the hypothalamus and orbitofrontal cortex change their firing rate to an identical food stimulus when the animal has been satiated on that particular food. From where does the orbitofrontal cortex get its inputs to convey information that the organism is satiated? There are brief paragraphs on somatic signals that would convey such information to the brain (p. 230, sections on gastric distension, chemosensation, glucose) but this is not developed. It has, however, been developed in some detail by others, for instance in the neuroanatomical reviews of Bud Craig (Craig, 2002), but this work is not cited anywhere in Rolls' book. There are several imaging studies and reviews on interoception also by one of Rolls' own students, Hugo Critchley, but these are not cited either (Critchley *et al.*, 2001).

The style of the book's prose is consistent with the reductionist philosophy of its author. Like Rolls' earlier book on emotion, *Emotion Explained* consists of a series of hierarchically structured and logically sequential chapters that articulate theory and data in simple sentences. Chapters typically begin with previews of the points to be made, often itemized, and conclude symmetrically. Numbered subsections enumerate specific issues. There is plentiful restatement of key points. The consequence is a volume that, given its comprehensive and ambitious nature, is probably the single substantial book on emotion that is the easiest to read.

While the title of the book leads one to assume that it is Professor Rolls' final word on the topic, I for one would right away buy a second edition, since it would be certain to provide a highly useful update on the literature. I would strongly recommend the book to any neuroscientist or psychologist interested in emotion, even to those who have already read *The Brain and Emotion*. It is not suitable for a lay readership, nor for those who do not care much about the brain. It should be required reading for all students in behavioural neuroscience, and has sufficient breadth that many of its chapters will be of interest also to experts in neurology, psychology or philosophy.

Ralph Adolphs
Division of Humanities and Social Sciences, HSS 228-77
California Institute of Technology
Pasadena, CA 91125, USA
E-mail: radolphs@hss.caltech.edu

doi:10.1093/brain/awl227

References

Bechara A, Damasio H, Tranel D, , Damasio A. The Iowa gambling task and the somatic marker hypothesis: some questions and answers. Trends Cogn Neurosci 2005; 9: 159–62.

Craig AD. How do you feel? Interoception: the sense of the physiological condition of the body. Nat Rev Neurosci 2002; 3: 655–66.

Critchley HD, Mathias CJ, Dolan RJ. Neuroanatomical basis for first- and second-order representations of bodily states. Nat Neurosci 2001; 4: 207–12.

Damasio AR. Descartes' error: emotion, reason, and the human brain. New York: Grosset/Putnam; 1994.

Glimcher PW. Decisions, uncertainty, and the brain: the science of neuroeconomics. Cambridge, MA: MIT Press; 2003.

Prinz J. Gut reactions. New York: Oxford University Press; 2004.

Rolls ET. The brain and emotion. New York: Oxford University Press; 1999.

Scherer KR. On the sequential nature of appraisal processes: indirect evidence from a recognition task. Cognition Emot 1999; 13: 763–94.