volumetric analysis based on three-dimensional reconstructions of magnetic resonance scans of human and ape brains *J. Hum. Evol.* 32, 375–388

46 Deacon, T.W. (1997) What makes the human brain different? *Annu. Rev. Anthropol.* 26, 337–357

47 Becker, L.E. *et al*. (1984) Dendritic development in human occipital cortical neurons *Brain Res.* 315, 117–124

48 Mrzljak, L. *et al*. (1990) Neuronal development in human prefrontal cortex in prenatal and postnatal stages *Prog. Brain Res.* 85, 185–222

49 Huttenlocher, P.R. (1994) Synaptogenesis, synapse elimination, and neural plasticity in human cerebral cortex, in *Threats to Optimal Development: Integrating Biological, Psychological, and Social Risk Factors* (*The Minnesota Symposia on Child Psychology, Vol. 27*) (Nelson, C., ed.), pp. 35–54, Erlbaum

50 Koenderink, M.J., Uylings, H.B. and Mrzljak, L. (1994) Postnatal maturation of the layer III pyramidal neurons in the human prefrontal cortex: a quantitative Golgi analysis *Brain Res.* 653, 173–182

51 Koenderink, M.J. and Uylings, H.B. (1995) Postnatal maturation of layer V pyramidal neurons in the human prefrontal cortex. A quantitative Golgi analysis *Brain Res.* 678, 233–243

52 Case, R. (1992) The role of the frontal lobes in the regulation of cognitive development (Special Issue: The role of frontal lobe maturation in cognitive and social development) *Brain Cognit.* 20, 51–73

53 Gibson, K.R. (1991) Myelination and behavioral development: a comparative perspective on questions of neoteny, altriciality and intelligence, in *Brain Maturation and Cognitive Development: Comparative and Cross-Cultural Perspectives* (*Foundations of Human Behavior*) (Gibson, K.R. and Petersen, A., eds), pp. 29–63, Aldine de Gruyter

54 Diamond, A. *et al*. (1997) Prefrontal cortex cognitive deficits in children treated early and continuously for PKU *Monogr. Soc. Res. Child. Dev.* 62

55 Goldman-Rakic, P.S. (1987) Development of cortical circuitry and cognitive function *Child Dev.* 58, 601–622

56 Luciana, M. and Nelson, C.A. (1998) The functional emergence of prefrontally guided working memory systems in four- to eight-year-old children *Neuropsychologia* 36, 273–293

57 Goldman-Rakic, P.S. (1995) Architecture of the prefrontal cortex and the central executive *Ann. New York Acad. Sci.* 769, 71–83

58 Fuster, J.M. (1995) Temporal processing *Ann. New York Acad. Sci.* 769, 173–181

59 Morris, R.G. *et al*. (1993) Neural correlates of planning ability: frontal lobe activation during the Tower of London test *Neuropsychologia* 31, 1367–1378

60 Baker, S.C. *et al*. (1996) Neural systems engaged by planning: a PET study of the Tower of London task *Neuropsychologia* 34, 515–526

61 Welsh, M.C., Pennington, B.F. and Groisser, D.B. (1991) A normative-developmental study of executive function: a window on prefrontal function in children *Dev. Neuropsychol.* 7, 131–149

62 Bachevalier, J. and Mishkin, M. (1984) An early and a late developing system for learning and retention in infant monkeys *Behav. Neurosci.* 98, 770–778

63 Schade, J.P. and van Groenigan, W.B. (1961) Structural organization of the human cerebral cortex: I. Maturation of the middle frontal gyrus *Acta. Anat.* 47, 72–111

Ten years of the rational analysis of cognition

Nick Chater and Mike Oaksford

Rational analysis is an empirical program that attempts to explain the function and purpose of cognitive processes. This article looks back on a decade of research outlining the rational analysis methodology and how the approach relates to other work in cognitive science. We illustrate rational analysis by considering how it has been applied to memory and reasoning. From the perspective of traditional cognitive science, the cognitive system can appear to be a rather arbitrary assortment of mechanisms with equally arbitrary limitations. In contrast, rational analysis views cognition as intricately adapted to its environment and to the problems it faces.

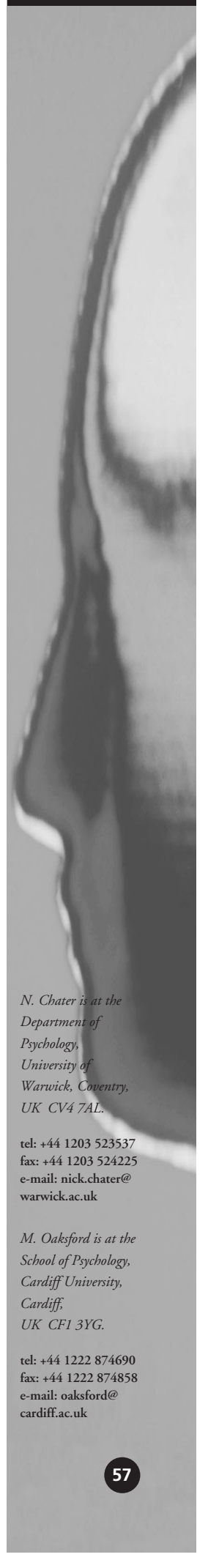
In 1989, J.R. Anderson and Milson[1] published the first paper explicitly adopting the 'rational analysis' approach to cognition. In the decade since, the approach has been vigorously pursued, whether by name[2–9] or merely in spirit[10–12]. Rational analysis has been the topic of an international conference, involving some of the world's leading cognitive psychologists and is the focus of the resulting book[13]. But what exactly is rational analysis? How does it relate to other approaches in cognitive science? How does it apply in practice? This review addresses these questions, beginning by distinguishing the style of explanation used in rational analysis from conventional explanation in the cognitive sciences.

What is rational analysis?

Mechanistic and purposive explanation

A scientific explanation of psychological, biological, or social phenomena can take one of two complementary forms. The first is 'mechanistic'. Phenomena are explained by analysing their internal causal structure. The second is 'purposive'. The phenomena are explained in terms of their purpose: what problem they solve.

In biology, purposive explanation concerns the *function* of biological structures and processes (e.g. the function of the heart is to pump blood); and the same style of explanation is applied to animal behaviour (e.g. the function of building nests is to provide a safe shelter for eggs). In social



N. Chater is at the Department of Psychology, University of Warwick, Coventry, UK CV4 7AL.

tel: +44 1203 523537
fax: +44 1203 524225
e-mail: nick.chater@warwick.ac.uk

M. Oaksford is at the School of Psychology, Cardiff University, Cardiff, UK CF1 3YG.

tel: +44 1222 874690
fax: +44 1222 874858
e-mail: oaksford@cardiff.ac.uk

1364-6613/99/$ – see front matter PII: S1364-6613(98)01273-X

Box 1. Rational analysis and evolutionary psychology

Rational analysis is concerned to explain why cognition is adaptive. A related approach is 'evolutionary psychology' (Ref. a). Whereas rational analysis is neutral concerning whether adaptation arises through evolution or learning, evolutionary psychology has been applied principally where adaptation arises through evolution. However, evolution might also have predisposed humans to learn some tasks more quickly than others.

Another point of contrast is that some evolutionary psychologists have argued that cognitive processes or behaviours that appear counteradaptive in the contemporary environment (to which adaptation might have occurred via learning) can be explained as having been adaptive in the hunter–gatherer societies of evolutionary history. Given that so little is known in detail about the problems faced in such societies, this style of explanation seems particularly weak. However, the general programme of evolutionary psychology is mainly concerned with how specific cognitive structures can be seen as adaptations to the environment. This emphasis on the environment in providing relevant goals for the cognitive system and crucial constraints on its operation is shared by the rational analysis approach (e.g. need-probability and the rarity assumption). Consequently, although rational analysis incorporates some elements of the evolutionary approach, some of its proponents (ourselves included) are less willing to endorse evolutionary 'just so' stories concerning primitive societies as explanations of cognitive phenomena.

This issue is highlighted in relation to the selection task, where some evolutionary psychologists have argued that the different performance observed when people are given so-called 'deontic' rules (i.e. norms of behaviour) can be explained by the evolutionary importance of reasoning about such social norms (Ref. b). With such rules, for example, 'if you drink beer (*p*), you must be over 18 (*q*)', people select the *p* and *not-q* cards, rather than the usual *p* and *q* cards selections. According to some evolutionary psychologists, this is a result of enhanced reasoning ability, owing to the operation of a dedicated social-reasoning module, wired in by evolutionary pressures on certain kinds of social interactions. But the task also has a different rational analysis, and hence a different 'correct' solution (Refs c,d), to the standard selection task, which explains the change of people's responses. Therefore, rational analysis explains the data parsimoniously, without requiring assumptions about evolutionary history [see the interchange between Cummins (Ref. e) and Chater and Oaksford (Ref. f) for further discussion].

References

a Barkow, J.H., Cosmides, L. and Tooby, J., eds (1995) *The Adapted Mind*, Oxford University Press

b Cosmides, L. (1989) The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task *Cognition* 31, 187–276

c Oaksford, M. and Chater, N. (1994) A rational analysis of the selection task as optimal data selection *Psychol. Rev.* 101, 608–631

d Oaksford, M. and Chater, N. (1995) Information gain explains relevance which explains the selection task *Cognition* 57, 97–108

e Cummins, D.D. (1996) Evidence for the innateness of deontic reasoning *Mind Lang.* 11, 160–190

f Chater, N. and Oaksford, M. (1996) Deontic reasoning, modules and innateness *Mind Lang.* 11, 191–202

theory, purposive explanation is embodied in 'rational choice' explanation[14], which ranges from economics to sociology and political science. People's behaviour is explained as rational in terms of their beliefs and goals (i.e. their purposes). This kind of explanation is a systematization of our everyday, folk-psychological explanation of each other's behaviour in terms of beliefs and desires.

In the cognitive sciences, however, there has been a strong predominance of mechanistic explanation. Computational models, whether symbolic or connectionist, have focused on specifying architectures and algorithms for cognition. Experimental studies have carefully documented the structural features of memory or reasoning, for example, but with relatively little concern for why these processes work as they do. Neuroscience provides another source of primarily mechanistic constraint on cognitive theory, in terms of the causal properties of the neural substrate in which cognitive processes are implemented. The picture that emerges from this focus on mechanistic explanation is of the cognitive system as an assortment of apparently arbitrary mechanisms, subject to equally capricious limitations, with no apparent rationale or purpose.

This rather unflattering picture of the cognitive system seems radically at variance with its performance. In perception, motor control, language processing, common-sense reasoning and decision making, the cognitive system reliably (though not infallibly) handles perceptual and cognitive problems of spectacular complexity, typically under conditions of extreme uncertainty. The cognitive system can learn to deal with a remarkably broad range of challenges, both natural and artificial, from unicycling to backgammon to musical composition. It acquires, stores, and can flexibly retrieve, an immensely rich understanding of the everyday world. It seems plausible that, as for other biological structures, this success is not accidental – rather, the cognitive system seems more likely to be superbly adapted to serve practical and computational ends.

Purposive explanation has been applied to some aspects of cognition. Most notably, there has been considerable interest in studying how visual processing is adapted to the problem of reconstructing environmental structure from visual input[15]. Marr's 'computational level' analysis aims to explain the nature of the problems that the visual system faces, and how these can be solved successfully[15]. The idea is that the adaptive success of visual processing can be explained by assuming that visual processes approximate an optimal solution to these problems. Another purposive style of explanation is the 'ideal observer' approach to understanding visual processes, in which the performance of a 'perfect' algorithm for solving a computational problem is compared with the visual system (which is assumed to attempt to solve that problem). This approach was developed in relation to quantal limits on light detection, but has recently been extended to complex visual functions, such as letter recognition[16], reading[11], and object recognition[17]. In the same vein, there has been much recent interest in how the visual system is adapted to the statistical properties of natural images[18].

The methodology of rational analysis

Although purposive explanation has been immensely fruitful in perception, it has been less vigorously pursued in cognition. This is presumably because analysing agents' goals and the environments in which they think and act appears

Box 2. The role of optimality

Optimization (Step 4 in a rational analysis) is not a straightforward step in carrying out a rational analysis. This is for at least two reasons. First, it might not be clear how to compute the optimal solution for a particular cognitive task. This issue separates rational analysis from the recent work of Chase and colleagues (Ref. a). On the one hand, according to rational analysis, deriving the optimal solution is the crucial step in explaining why human performance on a cognitive task is successful: it is successful to the extent that it *approximates* the optimal solution. Recently, for example, McKenzie (Ref. b) has shown that the linear heuristics thought to lead to irrational causal reasoning (Ref. c) actually provide very good approximations to the optimal solution. The need to explain the success of cognition means that, even if they are currently unavailable, deriving optimal solutions will remain a desideratum. Using Marr's analogy (Ref. d), ignoring this step of a rational analysis would be like trying to understand why birds can fly without a theory of aerodynamics. On the other hand, Gigerenzer and Goldstein (Ref. e), for example, seem to argue that environmentally successful algorithms can be developed without checking whether they approximate the optimal solution. However, Chase *et al.* (Ref. a) go on to analyse when their algorithms will be successful and when they will not, which is at least in the spirit of an optimality approach. Although rational analysis is committed to showing that cognitive algorithms approximate optimal solutions, whether this is necessary is an area of current debate.

Second, as we have seen, rational analysis follows explanations in economics and biology by focusing on optimal behaviour and assuming that actual behaviour approximates this. The tacit assumption is that good suboptimal behaviours will be similar to the optimal behaviour. But this is not necessarily true – in principle, it is possible that a problem could have two or more good solutions that are very different. Consider, for example, the range of classic algorithms in computer science for traveling to each of a set of locations covering the minimal possible distance (the Travelling-Salesman problem). There are many different good solutions, but often these solutions, although very close to the shortest path, will dictate very different paths both from the optimum path and from each other. If, given a particular set of locations, it was calculated that a particular path was optimal or near-optimal, it would therefore not at all follow that any good path would approximate this path. Putting the matter in slightly more general terms, many problems have deep 'local maxima' in the space of possible solutions; so good solutions corresponding to these maxima may be very distant from the global maximum (the optimal solution).

This adds a note of caution to the focus on optimality in rational analysis. Nonetheless, there are many aspects of cognition where this problem does not appear to arise in practice, including some of those outlined in this article.

References

a Chase, V.M., Hertwig, R. and Gigerenzer, G. (1998) Visions of rationality *Trends Cognit. Sci.* 2, 206–214

b McKenzie, C.R.M. (1994) The accuracy of intuitive judgment strategies: covariation assessment and Bayesian inference *Cognit. Psychol.* 26, 209–239

c Schustack, M.W. and Sternberg, R.J. (1981) Evaluation of evidence in causal inference *J. Exp. Psychol. Gen.* 110, 101–120

d Marr, D. (1982) *Vision*, W.H. Freeman

e Gigerenzer, G. and Goldstein, D.G. (1996) Reasoning the fast and frugal way: models of bounded rationality *Psychol. Rev.* 103, 650–669

more difficult to specify for higher-level cognition than it is for vision.

The program of rational analysis[2,3] provides a methodology for applying purposive explanation to higher cognitive processes. Rational analysis requires specification of the goals of the system and the nature of the environment, and formally derives an optimal solution to achieving some goal in that environment. Thus, rational analysis specifies what problem the cognitive system is solving – some approximation to an optimal solution. It also explains why this is useful: because it approximates the *optimal* solution to attaining some adaptively relevant goal or purpose (though it does not have a commitment to an evolutionary explanation for why the cognitive system is well-adapted to such goals; see Box 1).

Anderson's methodology[2,3] for deriving rational analyses of cognitive processes involves the following six steps.

(1) Goals: specify precisely the goals of the cognitive system.

(2) Environment: develop a formal model of the environment to which the system is adapted.

(3) Computational limitations: make minimal assumptions about computational limitations.

(4) Optimization: derive the optimal behaviour function, given 1–3 above.

(5) Data: examine the empirical evidence to see whether the predictions of the behaviour function are confirmed.

(6) Iteration: repeat, iteratively refining the theory.

Rational analysis is closely related to other approaches that have assumed that human behaviour is adapted to the structure of the environment. Psychologists such as Brunswick[19] and Gibson[20] adopted a specifically adaptionist perspective. Shepard[21] and Cosmides[22] have also pursued an adaptionist stance within an evolutionary framework (see Box 1). Recently Chase, Hertwig and Gigerenzer[23] have proposed a related approach that takes Herbert Simon's[24] notion of bounded rationality as its starting point. For Simon, 'Human rational behaviour…is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor'. Chase *et al.*[23] develop this analogy by proposing simple heuristics for human reasoning, which do not tax people's limited computational resources and which exploit specific features of the environment. Proponents of rational analysis adopt a similar perspective being equally concerned with modeling the structure of the environment (Step 2) and with computational limitations (Step 3) [the principal difference between these accounts concerns the role of optimization (Step 4) in cognitive explanation, which we discuss in Box 2]. We now illustrate how rational analysis applies in practice in two core cognitive domains – memory and reasoning.

Memory

A ubiquitous finding in memory research is that memory fails gradually over time. This is typically assumed to be a

Box 3. Need-probability and power functions

Anderson (Ref. a) assumes that an efficient memory system is one where the availability of a memory structure, S, is directly related to the probability that it will be needed. Need-probability, p, is a function of a history factor, H_S, and a context factor, $a(Qs)$:

$$p = P(H_S).a(Qs) \quad (1)$$

Where $P(H_S)$ is the probability that S is needed given its history of prior usage, and $a(Qs)$ represents the associations between a set of cues, Qs, that make up the current context and particular memory structures. Deriving an expression for the history factor presents the most problems.

Anderson (Ref. a) developed a theoretical model of the history factor based on previous analyses of informational retrieval systems (Ref. b). The problem is to derive an expression for the probability that a memory structure is needed given its history, $P(S|H_S)$. Two pieces of information about history are taken into account: the total time, t, that S has existed in memory and the total number of times, n, that S has been used. Previous models of information retrieval suggest prior distributions and likelihoods relevant to calculating $P(S|H_S)$ by Bayesian estimation (Ref. b). The expression for $P(S|H_S)$ is the mean of the posterior (Gamma) distribution:

$$P(S|H_s) = \frac{v+n}{M(t)+b} r(t), \text{where } M(t) = \int_0^t r(s)ds \quad (2)$$

$r(t)$ reflects the rate of decay in memory and v and b are the parameters of the prior distribution. If context is held constant and so need-probability depends only on the history factor, then Equation (2) predicts that need-probability (and hence the probability that a structure will be recalled) should be a negatively accelerating power function of retention interval, that is, $p = At^{-k}$.

A system operating according to Anderson's model will be rational in so far as need-probabilities are related to retention intervals as a power function. Only the structure of the environment determines whether this is the case: does the probability that someone will need to recall some information about, say, Saddam Hussein, vary as a power function of the time since they last heard anything about Saddam Hussein? Power functions are straight lines when plotted on log–log axes (Fig. I). Consequently, testing the predictions for environmental data involves examining the fit of the following equation to the data (reintroducing the context factor):

$$\log(p) = \log(A) - k\log(t) + \log[a(Qs)] \quad (3)$$

Equation (3) predicts graphs of the form shown in Fig. I. Exactly this form of relationship between retention interval and need-probability is reported by Schooler (Ref. c) in investigating environmental data about newspaper headlines. For example, the probability that Saddam Hussein will be mentioned in a headline in the *New York Times* varies as a power function of the number of days since he was last mentioned.

References

a Anderson, J.R. (1990) *The Adaptive Character of Thought*, Erlbaum

b Burrell, Q. (1980) A simple stochastic model for library loans *J. Document.* 36, 115–132

c Schooler, L. (1998) Sorting out core memory processes, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 128–155, Oxford University Press

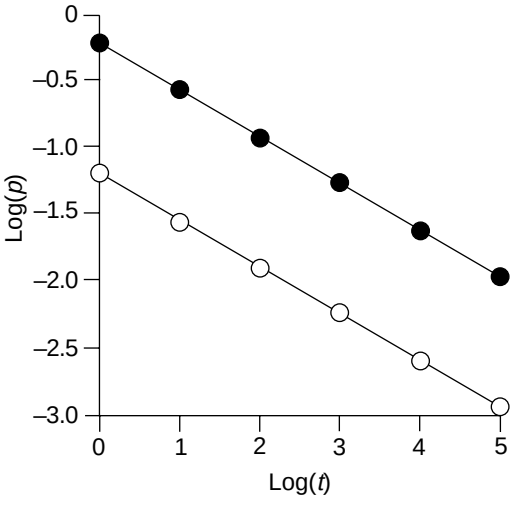


Fig. I. The relationship between log(*t*) and log(*p*) predicted by Equation (3) (see text). The slope is given by the exponent (k) of the power function and the intercept is given by log(A) + log[$a(Qs)$]. Two curves are shown, one where the cue associations are strong (filled circles) and one where they are weak (open circles). (The data shown are simulated.)

side-effect of problems of storage or retrieval in the memory system. But perhaps the pattern of memory breakdown can be viewed as adaptive. Perhaps recent items are remembered better because they are more likely to be needed again soon[1]. For example, if the last time you read a fact about Iraq was one sentence ago, then it is likely that Iraq will be mentioned in the next sentence – and that fact can be usefully drawn on. But if the last mention of Iraq was several days ago, then the probability that this information will be needed in understanding the next sentence is low. Perhaps rate of forgetting is optimally adapted to this decline in 'need-probability' over time. Let us consider the six steps of rational analysis as applied to memory in turn.

(1) Goals: the general goal of a memory system is to allow the efficient retrieval of relevant information from memory[1,2] (see Box 3). Specifically, this involves making the availability of a memory-trace match the probability that it will be needed.

(2) Environment: the environment determines the 'need-probability', p, for each item in memory, that is, the probability that some information will be needed, depending on its prior history of use. Roughly, items with high need-probability should be the most available in memory.

(3) Computational limitations: Anderson assumes that items in memory are searched sequentially, with a fixed cost, C, associated with searching each item in memory.

(4) Optimization: in general terms the optimization problem can be defined very simply. The memory system should stop retrieving memories when

$$pG < C$$

where G is the gain associated with retrieving a memory and C is the cost of searching for it. So attempts at retrieval

should stop when retrieval costs exceed the expected gain. The difficulty, which takes up most of Anderson's analysis, involves calculating a good estimate for p, which itelf depends on two factors: the current context (e.g. mentions of Saddam Hussein or Iraq) and an item's history of use (previous mentions of Saddam Hussein and Iraq). The latter presents the most problems. It is assumed that similar assumptions can be made about the usage of items in human memory as are made about other information retrieval systems, for instance, libraries[1,2]. The probability of an item being needed given its history is derived theoretically by extending existing models of usage in library systems[25].

(5) Data: the theoretical model predicts that need-probability is a decreasing power function of the time since an item was last studied (a power function has the form $y = x^a$) (see Box 3). Thus, simplifying Anderson's argument somewhat, if need-probability decays as a power function, then an adaptive memory system should forget items as a power function of time, a prediction that has been observed empirically[26]. Similarly, the model implies that need-probability increases as a power function of the frequency with which an item has been used in the past – this predicts the empirically ubiquitous power law of practice[27]. Finally, if number of exposures to an item, and time since last presentation are held constant, then there are interesting effects of spacing. If the final two presentations of an item are spaced at times $t-d$ (where d is a specific time interval) and t, then the need-probability is predicted to be maximal at $t+d$. This has been empirically observed[28,29].

(6) Iteration: Anderson and Schooler[30] have strengthened the empirical basis of their analysis of the memory 'environment' by looking at distributions of recurrence items in newspaper headlines and in parental speech. As in the theoretical model, need-probabilities estimated from these sources follow a power function.

On the empirical side, R.B. Anderson and colleagues have shown that by experimentally manipulating need-probability, forgetting functions can be manipulated[31]. Participants memorized strings of digits, and retained them for 1, 2, 4, 8 or 16 seconds. Some trials ended with a recall test, others with no test. Need-probability was manipulated by changing the proportion of tests for each retention interval (e.g. the probability of a test could increase or decrease with length of retention interval). Forgetting curves changed to reflect (although not exactly) these need probabilities. This is striking confirmation of the approach, which does not emerge from current mechanistic accounts of memory.

However, concerns have been raised about the fit between rational models and the presumed power law of forgetting[32]. Apparently, such power laws can be artifacts arising from averaging non-power-law forgetting curves across individual subjects. However, it has been countered that individual forgetting curves also follow a power law, in line with J. Anderson's account[33]. Nonetheless, the power law of forgetting remains a centre of controversy[34,35].

Schooler[9] has extended empirical tests to deal with both of the factors assumed to determine need-probability: the context in which an item occurs and its history of use. Environmental analysis of the joint effects of time elapsed and the presence of associated items leads to precise predictions from the rational analysis concerning the interaction of these factors in memory. Schooler tested these predictions using a cued-recall task in which the cues were either strongly associated or not associated with the targets, using various time delays. He found that the interaction of time and contexts was broadly in line with predictions of the rational analysis. But there were interesting disparities with the model; for example, that time elapsed was relatively more important than contextual cues than would be expected from an optimal match with the environment. Schooler argues that this can be explained in terms of the underlying processing mechanisms that are assumed to instantiate the rational analysis. Similarly, R. Anderson[36] argues that processing mechanisms (e.g. the capacity of short-term memory) must be taken into account to explain the departure from the prediction of the rational analysis – this requires a richer specification at Step 3 of the rational analysis.

A range of other recent work also focuses on the interface between rational- and algorithmic-level explanations of memory[37,38]. This emphasis on welding together rational analysis with cognitive mechanisms also underlies the latest version of the Anderson's ACT cognitive architecture, ACT-R (for 'rational'), which incorporates aspects of the analysis of memory directly into the computational architecture[5]. The interaction between the rational and mechanistic explanation of memory is likely to be an important area of future research.

Reasoning

In the cognitive science of reasoning, approaches based on rational analysis have been developed in a number of areas. Indeed, in the study of reasoning there is some research implicitly adopting the rational analysis approach[39] that predates Anderson and Milson's[1] paper. We focus on how rational analysis applies to the most controversial and heavily researched reasoning task, Wason's selection task[40], and then touch briefly on other important developments.

In the abstract selection task, people are given a rule of the form *if p then q* (e.g. 'if there is an A on one side there is a 2 on the other') and are presented with four cards, each of which has a p (e.g. A) or a *not-p* (e.g. K) on one side, and a q (e.g. 2) or *not-q* (e.g. 7) on the other side. Only the upper faces of the cards are visible, and these faces are as shown in Fig. 1. The task is to decide which cards need to be turned over in order to decide whether the rule is true or false.

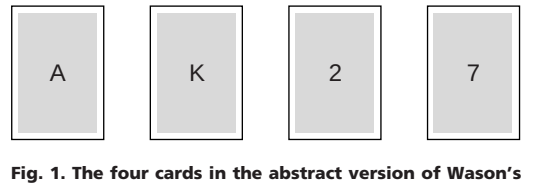


Fig. 1. The four cards in the abstract version of Wason's selection task. Participants are told that each card has a number on one side and a letter on the other side. They are also given a rule, such as, 'if there is an A on one side, then there is a 2 on the other side'. Participants are then asked to select the cards they 'must' turn over in order to decide whether the rule is true or false.

Box 4. Expected information gain

According to the optimal data selection (ODS) model, people are initially uncertain whether a rule is true. Thus they assign equal probabilities to the hypothesis (H_D) that the rule, *if p then q*, is true and the hypothesis that *p* and *q* are independent (H_I), that is, $P(H_D) = P(H_I) = 0.5$. Using standard information theory (Ref. a), this means that the uncertainty about whether H_D is true, $I(H_D)$, is maximal at 1 bit. In order to determine how conducting an experiment (*e*), (e.g. turning a card in the Wason selection task; see Fig. 1, main article) reduces this uncertainty it is necessary to calculate the probability that H_D is true given the data *D*, $P(H_D|D, e)$, by Bayesian estimation. For example, what reduction in uncertainty is obtained by conducting the experiment of turning the 2 (*q*) card to reveal some data (*D*), that is, either an A (*p*) or a K (*not-p*)? Applying Bayes' theorem requires information about the likelihoods of these outcomes given this experiment under the different possible hypotheses, that is, $P(\text{A}|2,H_D)$, $P(\text{K}|2,H_D)$, $P(\text{A}|2,H_I)$, and $P(\text{K}|2,H_I)$. These likelihoods can be calculated directly from the contingency tables used to represent H_D and H_I in the ODS model. So the probability that H_D is true given that, for example, the 2 card is turned to reveal an A [$P(H_D|\text{A}, 2)$], can be calculated by Bayes' theorem. This value can be used to calculate the uncertainty remaining after selecting the 2 card to reveal an A, that is, $I(H_D|\text{A}, 2)$. However, in the selection task, participants can not actually turn the cards over to reveal the data. Therefore the *expected* value of uncertainty after turning the 2 card, $E[I(H_D|2)]$, must be calculated instead. So $I(H_D|\text{A}, 2)$ and $I(H_D|\text{K}, 2)$ are weighted by the expected values of $P(\text{A}|2)$ and $P(\text{K}|2)$, respectively. The difference between the initial uncertainty, $I(H_D)$, and the expected uncertainty after turning a card, $E[I(H_D|2)]$, provides the 'information gain' that can be expected from conducting the experiment (*e*) of turning the 2 card. More generally, the expected information gain associated with conducting experiment *e*, $EI_g(e)$, is given by:

$$EI_g(e) = I(H_D) - E[I(H_D|e)]$$

Reference

a Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*, University of Illinois Press

Logically, participants should select only the *p* and the *not-q* cards, that is, those cards with the potential to reveal a falsifying instance. However, as few as 4% of participants make this response, other responses being far more common [*p* and *q* cards (46%); *p* card only (33%), *p*, *q* and *not-q* cards (7%), *p* and *not-q* cards (4%)][41].

Viewed abstractly, the selection task involves optimal data selection, as discussed in statistics and the philosophy of science. There is a hypothesis (the rule) and the problem is to decide which experiments should be conducted (cards should be turned) in order to help decide whether the hypothesis is true. Different views of the philosophy of science give different recommendations concerning how this problem should be solved.

According to Popper's falsificationist view of science[42], it is never possible to confirm a rule, only to disconfirm it. The rule will be disconfirmed only if there is a logical contradiction between the hypothesis and the data; that is, this requires finding a *p*, *not-q* card, which is explicitly disallowed by the rule. Hence the *p* card should be chosen, because it might have a *not-q* on the back, and the *not-q* card should be chosen, because it might have a *p* on the back, and no other cards should be chosen. This is the 'logical' solution to the task (although there are contrasting views[43]). The fact that people do not follow this pattern has been taken as casting doubt on human rationality[44–46].

But considering an everyday example reveals that this recommendation is counterintuitive. Suppose that the rule is: 'if a saucepan falls (*p*), it clangs (*q*)'. If I see the saucepan fall (*p* card), I should listen for a clang (*q* card) (e.g. take off my headphones). If there is a clang, the rule seems more plausible; if there is not, the rule appears disconfirmed. Similarly, if I hear a clang while sitting in the next room (*q*), I should look to see whether a saucepan has fallen – if it has, the rule seems more plausible. But, if I hear no clang while sitting in the next room, it seems futile to look to see if a saucepan has fallen. It is *possible* that this has happened, in which case the rule would be disconfirmed; but as saucepans fall rather rarely, it is more likely that no saucepan has fallen and I will learn nothing. In this example, the *q* card seems more worth turning than the *not-q* card – contrary to the 'logical' solution, but in line with performance on the selection task. This example suggests that in an environment where most properties and events are rare, looking for apparently confirmatory evidence will be more informative. The assumption that by default people treat properties in the environment as rare is central to providing a rational analysis of normal selection-task performance.

Oaksford and Chater's[7] rational analysis draws on the theory of optimal data selection (ODS) from Bayesian statistics[47], rather than on Popper's falsificationism. The six steps of the ODS model are as follows:

(1) Goals: selecting the data that has the greatest expected informativeness (EI_g) about whether the rule is true or whether the antecedent (*p*) and consequent (*q*) of the rule are independent (see Box 4).

(2) Environment: the ODS model relies crucially on the assumption that properties (e.g. falling; clanging) are rare (i.e. most things are not falling or clanging at most times). Given this 'rarity' assumption it makes sense to check if the saucepan has fallen given the rare event of hearing a clang (*q* card), but not to check given the very common occurrence of hearing no clang (*not-q* card). This intuition is captured by the ODS model. When properties are rare, that is, $P(p)$ and $P(q)$ are low [$P(p) < 0.4$ and $P(q) < 0.25$ approximately] then $EI_g(q) > EI_g(\textit{not-q})$.

(3) Computational limitations: it is assumed that there is a cost of examining data, so as little as possible is examined.

(4) Optimization: the ODS model assumes that people select the most informative data they can, subject to the costs of turning the cards that they are willing to accept. Assuming rarity leads to the following order of expected information gain across the four cards: $EI_g(p) > EI_g(q) > EI_g(\textit{not-q}) > EI_g(\textit{not-p})$. Costs determine the number of

cards that a participant will turn over. Thus, maximizing information gain involves turning cards in the above order up to this number. So, if participants decided to turn just one card, this would be the most informative single card, *p*; if they decided to turn two cards, they would turn the two most informative cards, *p* and *q*; and if they decided to turn three cards, they would turn the three most informative cards, *p*, *q* and *not-q*.

(5) Data: the ODS model accounts for the large volume of apparently puzzling data on Wason's selection task. It resolves the problem of apparent human irrationality, by demonstrating that human performance approximates Bayesian optimal data selection (see Fig. 2). This model captures all the main experimental results on the selection task, including the non-independence of card selections[48], the negations paradigm[49], the therapy experiments[50], the reduced-array selection task[51], and work on so-called fictional outcomes[52].

(6) Iteration: the ODS model makes the novel prediction that selection-task performance should change if the rarity assumption is violated. This has been tested and observed empirically[53–55]. As in the case of memory, this confirmation is striking, because it would not have been predicted by mechanistic models.

Other probabilistic models of the selection task have been proposed, building on the ODS model[56–59]. Chater and Oaksford[60] have also developed a related account of syllogistic reasoning, which aims to model even this paradigmatic area of logical inference using a probabilistic rational explanation.

Another area of reasoning research where rational explanation has led to a radical re-interpretation of apparently flawed human inference is causal reasoning. For example, Cheng[10] has developed the 'Power PC theory' of causal reasoning. Roughly speaking, in this model the main determinant of the causal influence of a factor on some outcome is the difference in probability of the outcome when the causal factor is, or is not, present, if all other factors are held constant. This model can be viewed as a rational analysis for the Rescorla–Wagner model of classical conditioning[12]; and moreover, it reveals that apparent 'biases' in causal inference[61] are predicted by a rational model, and hence are not really biases at all.

Interestingly, there have recently been attempts to integrate the two themes we have discussed, by considering causal conditions in the selection task. Over and Jessop[62] argue that apparent biases in causal judgments using 2×2 contingency tables correspond to apparent biases in selection-task performance. Consequently if selection-task performance can be viewed as rational, then so can causal judgments, using directly analogous arguments.

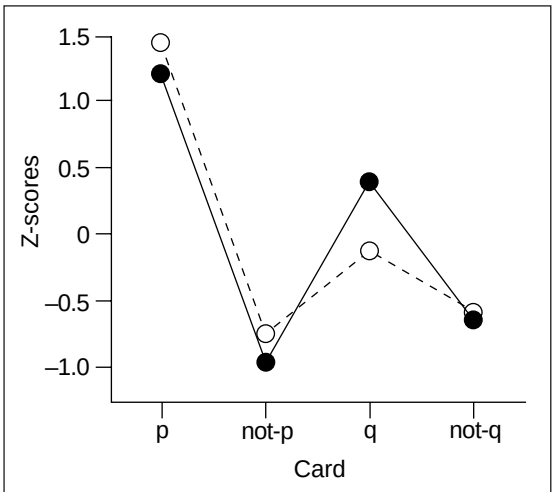


Fig. 2. Comparison of the expected information gain (dotted line) with the frequency of card selections in the standard abstract selection task (solid line). (For purposes of comparison both scales have been normalized.) These data are taken from the meta-analysis of the selection task reported by Oaksford and Chater[7].

Conclusion

Traditional cognitive psychology implicitly treats the cognitive system as a ragbag of arbitrary mechanisms, with arbitrary performance limitations. Little attention is given to why these mechanisms and limitations add up to a system that is so adaptively successful in coping with a complex and partially known world. Rational analysis aims to answer this question by identifying the problems that specific cognitive mechanisms face, and crucially by including the environment in which these problems occur.

Working out 'optimal behaviour functions' – optimal solutions for solving these problems – provides the possibility of explaining why cognitive mechanisms are as they are (although see Box 2); it also provides a major source of constraint on the development of theories of specific cognitive processes, and, as we have seen, is a source of novel empirical predictions. Rational analysis has been remarkably fruitful in the key domains of memory and reasoning. We suggest that it could play a central role in the development of cognitive science as a whole.

Outstanding questions

- What are the limits of rational analysis? Can every successful aspect of cognition be explained rationally, or could some mechanisms 'just work' with no rational explanation[63]?
- How far can rational analysis in cognitive science be integrated with related work in perception and motor control, such as Marr's[15] 'computational level' of explanation, 'ideal-observer' models'[64], and 'task dynamics'[65]?
- How does rational analysis relate to proposed cognitive architectures? We have seen that Anderson's ACT-R architecture is designed to have close links with aspects of rational analysis. Many connectionist networks also have probabilistic interpretations, which may be viewed as rational analyses for explaining why and when these networks learn and behave effectively[66,67].
- Can learning be given a rational analysis? Formal research on computational theories of learning; for example, based on Kolmogorov complexity[68], the Vapnik–Chervonenkis dimension[69], or standard Bayesian statistics, could provide a general framework for such rational analyses.
- How constrained is rational analysis? One of the problems with theorizing about the algorithms that the cognitive system might use is that there are typically many different plausible algorithms that are consistent with the observed data. This was one of Anderson's[2] motivations for developing the rational analysis approach. But is rational analysis really any more constrained? Or are there always many plausible rational analyses compatible with the data?

Acknowledgements

We would like to thank John R. Anderson and Gerd Gigerenzer for discussion of these ideas. We also thank David Over and the anonymous reviewers for their extremely helpful comments on this article. Thanks especially to Jean Czerlinski for further discussions.

References

1 Anderson, J.R. and Milson, R. (1989) Human memory: an adaptive perspective *Psychol. Rev.* 96, 703–719

2 Anderson, J.R. (1990) *The Adaptive Character Of Thought*, Erlbaum

3 Anderson, J.R. (1991) Is human cognition adaptive? *Behav. Brain Sci.* 14, 471–517

4 Anderson, J.R. (1991) The adaptive nature of human categorization *Psychol. Rev.* 98, 409–429

5 Anderson, J.R. (1994) *Rules of The Mind*, Erlbaum

6 Nosofsky, R. (1991) Relation between the rational model and the context model of categorization *Psychol. Sci.* 2, 416–421

7 Oaksford, M. and Chater, N. (1994) A rational analysis of the selection task as optimal data selection *Psychol. Rev.* 101, 608–631

8 Oaksford, M. and Chater, N. (1998) *Rationality in an Uncertain World*, Psychology Press

9 Schooler, L. (1998) Sorting out core memory processes, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 128–155, Oxford University Press

10 Cheng, P.W. (1997) From covariation to causation: a causal power theory *Psychol. Rev.* 104, 367–405

11 Legge, G.E., Klitz, T.S. and Tjan, B.S. (1997) Mr Chips: an ideal-observer model of reading *Psychol. Rev.* 104, 524–553

12 Shanks, D.R. (1995) Is human learning rational? *Q. J. Exp. Psychol.* 48A, 257–279

13 Oaksford, M. and Chater, N., eds (1998) *Rational Models of Cognition*, Oxford University Press

14 Elster, J., ed. (1986) *Rational Choice*, Blackwell Science

15 Marr, D. (1982) *Vision*, W.H. Freeman

16 Parish, D.H. and Sperling, G. (1991) Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination *Vis. Res.* 31, 1399–1415

17 Liu, Z., Knill, D.C. and Kersten, D. (1995) Object classification for human and ideal observers *Vis. Res.* 35, 549–568

18 Pelah, A. (1997) The vision of natural and complex images *Vis. Res.* 37, 3201–3439

19 Brunswick, E. (1956) *Perception and the Representative Design of Psychological Experiments* (2nd edn), University of California Press

20 Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Houghton-Mifflin, Boston

21 Shepard, R.N. (1984) Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, and dreaming *Psychol. Rev.* 91, 417–456

22 Cosmides, L. (1989) The logic of social exchange: has natural selection shaped how humans reason: studies with the Wason selection task *Cognition* 31, 187–276

23 Chase, V.M., Hertwig, R. and Gigerenzer, G. (1998) Visions of rationality *Trends Cognit. Sci.* 2, 206–214

24 Simon, H. (1990) Invariants in human behavior *Annu. Rev. Psychol.* 41, 1–19

25 Burrell, Q. (1980) A simple stochastic model for library loans *J. Document.* 36, 115–132

26 Wickelgren, W.A. (1976) Memory storage dynamics, in *Handbook of Learning and Cognitive Processes* (Estes, W.K., ed.), Erlbaum

27 Newell, A. and Rosenbloom, P. (1981) Mechanisms of skill acquisition and the law of practice, in *Cognitive Skills and Their Acquisition* (Anderson, J.R., ed.), pp. 1–55, Erlbaum

28 Bahrick, H.P. (1979) Maintainance of knowledge: questions about memory we forget to ask *J. Exp. Psychol. Gen.* 108, 296–308

29 Glenberg, A.M. (1976) Monotonic and nonmonotonic lag effects in paired-associated and recognition memory paradigms *J. Verb. Learn. Verb. Behav.* 15, 1–16

30 Anderson, J.R. and Schooler, L.J. (1991) Reflections of the environment in memory *Psychol. Sci.* 2, 396–408

31 Anderson, R.B. *et al.* (1997) Need-probability affects retention: a direct demonstration *Mem. Cognit.* 25, 867–872

32 Anderson, R.B. and Tweney, R.D. (1997) Artifactual power curves *Mem. Cognit.* 25, 724–730

33 Wixted, J.T. and Ebbesen, E.B. (1997) Genuine power curves in forgetting: a quantitative analysis of individual subject forgetting functions *Mem. Cognit.* 25, 731–739

34 Wixted, J.T. and Ebbesen, E.B. (1991) On the form of forgetting *Psychol. Sci.* 2, 409–415

35 Rubin, D.C. (1996) One hundred years of forgetting: a quantitative description of retention *Psychol. Rev.* 103, 734–760

36 Anderson, R.B. (1998) Rational and non-rational aspects of forgetting, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 156–164, Oxford University Press

37 Dennis, S. and Humphreys, M. (1998) Cueing for context: an alternative to global models of recognition memory, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 109–129, Oxford University Press

38 Shiffrin, R.M. and Steyvers, M. (1998) The effectiveness of retrieval from memory, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 73–95, Oxford University Press

39 Klayman, J. and Ha, Y. (1987) Confirmation, disconfirmation and information in hypothesis testing *Psychol. Rev.* 94, 211–228

40 Wason, P.C. (1968) Reasoning about a rule *Q. J. of Exp. Psychol.* 20, 273–281

41 Johnson-Laird, P.N. and Wason, P.C. (1970) A theoretical insight into a reasoning task *Cognit. Psychol.* 1, 134–148

42 Popper, K.R. (1959) *The Logic of Scientific Discovery*, Hutchinson

43 Sperber, D., Cara, F. and Girotto, V. (1995) Relevance theory explains the selection task *Cognition* 57, 31–95

44 Cohen, L.J. (1981) Can human irrationality be experimentally demonstrated? *Behav. Brain Sci.* 4, 317–370

45 Stein, E. (1996) *Without Good Reason*, Oxford University Press

46 Stich, S. (1990) *The Fragmentation of Reason*, MIT Press

47 Lindley, D.V. (1956) On a measure of the information provided by an experiment *Ann. Math. Stats.* 27, 986–1005

48 Pollard, P. (1985) Nonindependence of selections on the Wason selection task *Bull. Psychonomic Soc.* 23, 317–320

49 Evans, J.St-B.T. and Lynch, J.S. (1973) Matching bias in the selection task *Br. J. Psychol.* 64, 391–397

50 Wason, P.C. (1969) Regression in reasoning *Br. J. Psychol.* 60, 471–480

51 Johnson-Laird, P.N. and Wason, P.C. (1970) Insight into a logical relation *Q. J. Exp. Psychol.* 22, 49–61

52 Kirby, K.N. (1994) Probabilities and utilities of fictional outcomes in Wason's four-card selection task *Cognition* 51, 1–28

53 Green, D.W., Over, D.E. and Pyne, R.A. (1997) Probability and choice in the selection task *Think. Reason.* 3, 209–235

54 Manktelow, K.I., Sutherland, E.J. and Over, D.E. (1995) Probabilistic factors in deontic reasoning *Think. Reason.* 1, 201–220

55 Oaksford, M. *et al.* (1997) Optimal data selection in the reduced array selection task (RAST) *J. Exp. Psychol. Learn. Mem. Cognit.* 23, 441–458

56 Evans, J.St-B.T. and Over, D.E. (1996) Rationality in the selection task: epistemic utility versus uncertainty reduction *Psychol. Rev.* 103, 356–363

57 Klauer, K.C. On the normative justification for information gain in Wason's selection task *Psychol. Rev.* (in press)

58 Oaksford, M. and Chater, N. (1996) Rational explanation of the selection task *Psychol. Rev.* 103, 381–391

59 Chater, N. and Oaksford, M. Information gain vs. decision-theoretic approaches to data selection: Response to Klauer *Psychol. Rev.* (in press)

60 Chater, N. and Oaksford, M. The probability heuristics model of syllogistic reasoning *Cognit. Psychol.* (in press)

61 Schustack, M.W. and Sternberg, R.J. (1981). Evaluation of evidence in causal inference *J. Exp. Psychol. Gen*. 110, 101–120

62 Over, D.E. and Jessop, A. (1998) Rational analysis of causal conditionals and the selection task, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 399–414, Oxford University Press

63 Gigerenzer, G. and Goldstein, D.G. (1996) Reasoning the fast and frugal way: models of bounded rationality *Psychol. Rev.* 103, 650–669

64 Geisler, W.S. (1989) Sequential ideal-observer analysis of visual discriminations *Psychol. Rev.* 96, 267–314
65 Saltzman, E. and Kelso, J.S. (1987) Skilled actions: a task-dynamic approach *Psychol. Rev.* 94, 84–106
66 Chater, N. (1995) Neural networks: the new statistical models of mind, in *Connectionist Models of Memory and Language* (Levy, J.P. *et al.*, eds), pp. 207–227, UCL Press
67 McClelland, J.L. (1998) Connectionist models and Bayesian inference, in *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 21–53, Oxford University Press
68 Li, M. and Vitanyi, P. (1997) *An Introduction to Kolmogorov Complexity Theory and its Applications*, Springer-Verlag
69 Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*, Springer-Verlag

Review

Amygdala circuitry in attentional and representational processes

Peter C. Holland and Michela Gallagher

The amygdala has long been implicated in the display of emotional behavior and emotional information processing, especially in the context of aversive events. In this review, we discuss recent evidence that links the amygdala to several aspects of food-motivated associative learning, including functions often characterized as attention, reinforcement and representation. Each of these functions depends on the operation of separate amygdalar subsystems, through their connections with other brain systems. Notably, very different processing systems seem to be mediated by the central nucleus and basolateral amygdala, subregions of the amygdala that differ in their anatomy and in their connectivity. The basolateral amygdala is involved in the acquisition and representation of reinforcement value, apparently through its connections with ventral striatal dopamine systems and with the orbitofrontal cortex. The dentral nucleus, however, contributes heavily to attentional function in conditioning, by way of its influence on basal forebrain cholinergic systems and on the dorsolateral striatum.

The view that the amygdala and other temporal lobe structures are critical to the display of emotional behavior has prevailed for more than 60 years. For example, Kluver and Bucy[1] reported that monkeys with temporal lobe lesions became less prone to display fear, disgust, and other emotional behavior. Similarly, Kaada[2] described organized fear and rage responses in cats in response to electrical stimulation of the amygdala.

In an early review, Weiskrantz[3] concluded that amygdala function might be described as the attachment of emotional significance to stimuli by stimulus–reward learning. Since that pivotal report, many researchers have focused on the role of the amygdala in learning, culminating in elegant descriptions of the contributions of the amygdala to fear conditioning by Davis, LeDoux, and others[4–6]. These investigations have done for the amygdala and fear conditioning what the work of Thompson[7] and others has done for the cerebellum and simple motor conditioning, such as the eye blink response. These two systems are probably the most studied and best defined mammalian model systems for the study of the neurobiology of associative learning.

In the wake of this success in the analysis of amygdala functions in fear conditioning, it is easy to lose sight of the fact that the amygdala is busy performing other tasks as well. Some of these other functions fit under the general rubric of emotion or affect, and others do not. In the same way that decades of research in motor control shows that the cerebellum's job is not limited to the blinking of an eye, considerable evidence shows that the amygdala participates in many adaptive behavioral functions in addition to fear conditioning. Furthermore, some of this evidence suggests that the organization of information processing in these functions differs from that portrayed in the flow charts typically proposed for simple fear conditioning (Box 1).

In this article, we discuss recent evidence implicating the amygdala in several aspects of associative learning apart

P.C. Holland is at the Department of Psychology: Experimental, Duke University, Durham, NC 27708-0086, USA.
M. Gallagher is at the Department of Psychology, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218-2686, USA.

tel: +1 919 660 5699
fax: +1 919 660 5726
e-mail: pch@duke.edu

1364-6613/99/$ – see front matter PII: S1364-6613(98)01271-6