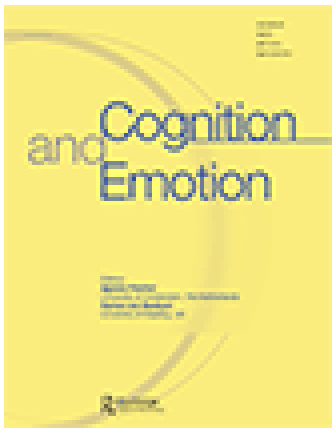

Cognition and Emotion
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pcem20

Towards a Cognitive Theory of Emotions

Keith Oatley [a] & P. N. Johnson-laird [b]
[a] University of Glasgow , Glasgow, UK
[b] MRC Applied Psychology Unit , Cambridge, UK
Published online: 07 Jan 2008.

To cite this article: Keith Oatley & P. N. Johnson-laird (1987) Towards a Cognitive Theory of Emotions, Cognition and Emotion, 1:1, 29-50, DOI: 10.1080/02699938708408362

To link to this article: http://dx.doi.org/10.1080/02699938708408362

COGNITION AND EMOTION, 1987, *1* (1) 29–50

Towards a Cognitive Theory of Emotions

Keith Oatley
University of Glasgow, Glasgow, U.K.
P. N. Johnson-Laird
MRC Applied Psychology Unit, Cambridge, U.K.

A theory is proposed that emotions are cognitively based states which co-ordinate quasi-autonomous processes in the nervous system. Emotions provide a biological solution to certain problems of transition between plans, in systems with multiple goals. Their function is to accomplish and maintain these transitions, and to communicate them to ourselves and others. Transitions occur at significant junctures of plans when the evaluation of success in a plan changes. Complex emotions are derived from a small number of basic emotions and arise at junctures of social plans.

INTRODUCTION

In the first study of emotions to be based on the theory of evolution, Darwin (1872) concluded that emotional expressions are a kind of neural accident. They result from overflows of neural excitation which may serve no function in the actions of adults. Had he lived later, Darwin might have used as an example of this superfluity the facial expressions that people make when speaking on the telephone. His theory was that emotional expressions are vestiges of evolutionary history or of childhood habit.

This idea that major aspects of emotions are not functional has been taken up by later writers, who include Dewey (1895) in his attempt to reconcile Darwin's work with James' (1890). He proposed that emotions are disturbances which occur when habits are no longer useful in present contexts.

Requests for reprints should be sent to Professor K. Oatley, Department of Psychology, University of Glasgow, Glasgow G12 8RT, U.K. This paper is based on a presentation to the Summer Institute on Cognition–Emotion Interrelations held in Colorado, 18–23 August, 1985, and supported by NIMH. We very gratefully acknowledge helpful discussions with Steven Draper, George Mandler and Richard Power, all of whom commented on a draft of the paper.

Dewey's proposal was an example of what Mandler (1984) has called a conflict theory of emotion. In this type of theory, and putting the problem in cognitive terms, emotions arise as disturbances which accompany interruptions and discrepancies among multiple goals and representations. The idea has been developed by cognitive and computational theorists (Mandler, 1964; Simon, 1967; Sloman and Croucher, 1981; Sloman, in press).

Our theory is a conflict theory in Mandler's sense. Deriving from the version of conflict theory that Simon (1967) proposed, our theory also proposes that emotions have important cognitive functions. Emotions may be disturbing, but they are not just inco-ordinations or side effects.

In this paper we use the term "cognitive" to refer to psychological explanations in terms of the representation and transformation of knowledge which may or may not be conscious. Tacit knowledge is used, for instance, when a cuckoo, hatched in Britain, leaves at the end of the summer to fly to Africa without benefit of parental training; or when a human speaker of a language utters sentences grammatically although unable to give an explicit account of that language's syntax.

In this cognitive approach, we refer to goals as symbolic representations of possible states of the environment that a system will try to achieve. Plans are sequences of transformations between representations, that link a current state of the environment to a goal. To make a plan is to assemble such a sequence. To carry it out is to enact the sequence in the world. We use the terminology of goals and plans to indicate aims and assembled sequences of action in a way that is neutral as to whether they are conscious. Thus a conscious plan might be an arrangement to catch a particular train; plans which are at least partly unconscious include both those that are instinctive and those that are highly practiced like tying a shoelace.

We will also use the notion of evaluation of a plan in a neutral way. If an evaluation is conscious we will use the word "conscious", which will here not be a synonym of "cognitive".

The theory is not a contribution to the debate about whether emotions require conscious thought, which has been stimulated by Zajonc's (1980) argument that preferences need no inferences. We do not claim that all emotions derive from thinking. Some do and some do not.

The purpose of this approach and this terminology is to lay out a groundwork of emotional theory that relates to the computationally based theories of language and perception in cognitive science (see e.g. Johnson-Laird, 1983a; Gardner, 1985) in which emotions have been very neglected, although, as we argue, they are central to the organisation of cognitive processing. The approach is intended to be a step towards theories of emotions that can be tested formally, and their implications explored in computer programs.

As part of the argument that emotions have important cognitive functions, we propose that they are part of a management system to co-ordinate each individual's multiple plans and goals under constraints of time and other limited resources. We go beyond Simon's (1967) suggestions along these lines, to propose a specific system of internal communication to achieve this management, and we also propose that emotions are further evolved in social species such as ourselves to communicate junctures in mutual plans among individuals in social groups. The human skin is, as it were, permeable to emotions: an emotion such as anxiety is not only communicated through the individual human body and brain to set them into a particular state appropriate to danger, but it also propagates beyond the individual to influence others and may set them into a similar or complementary state.

What should a theory of the emotions explain? The basic observation is that adult human beings report subjective experiences with a particular phenomenological tone that they describe as emotional. These experiences are consciously preoccupying. They are typically accompanied by certain somatic events, revealed by facial and other expressions, and lead to certain characteristic courses of action. A theory should account for these components, for the diversity of emotions, for the variation in their quality, and for their relations with other aspects of mental life. It should also answer the question of whether emotions have a function.

In our theory we bring together and develop two threads of theorising: one concerns the function of emotions in modular nervous systems, the other describes the occurrence of emotions at significant junctures of plans influenced by multiple goals.

SCHEDULING PROCESSES IN A MODULAR COGNITIVE SYSTEM

Hierarchical Organisation of Processors

Basic to our theory is the proposal that the human cognitive system is modular and asynchronous as described by Johnson-Laird (1983a, 1983b). Johnson-Laird describes a module as an autonomous processor. It may be thought of as a procedure that is relatively self-contained, and that attempts to complete its computation once it is activated. Each module has an associated goal, and can accomplish it given certain preconditions and absence of interruptions. In such systems problems arise as to how to co-ordinate processors that only compute when they receive the right input; how, for instance, to avoid pathological situations in which two processors are each awaiting an input from the other. One solution is to build hierarchies in which processors at higher levels invoke lower proces-

sors, in the way described by Miller, Galanter and Pribram (1960). This method is familiar both from neurological theorising and from computer programs with many levels of embedding in which one procedure can call others as subroutines, these in turn can call sub-sub-routines, and so on.

At the top level of the hierarchy of modules in the human cognitive system there is a processor corresponding to an operating system, capable of invoking lower level processors in specific sequences or according to particular pattern matches. The operating system needs to include, as Minsky (1968) implies, some model of the whole system. That is, the human cognitive system needs to include a model of itself, though the implications of this idea have as yet to be worked out in cognitive science.

The core of our proposal is that emotions are based on one of two specific kinds of communication in such systems. One kind of communication is propositional. Propositional signals are symbolic: i.e. they have internal structure that plays a part in denotation within the system. These signals correspond to calling patterns and procedure names that pass down the hierarchy to invoke lower level functions, to representations of different aspects of the world, to results and arguments of functions, and to messages that can construct new procedures.

The other kind of communication is non-propositional. It is simpler, cruder, and evolutionarily older. Non-propositional signals have no internal symbolic structure of significance to the system. They do not denote anything. Like hormones, they function purely causally. They propagate globally among the processors to set them into specific modes at particular junctures of multi-goal planning sequences. Emotions are based on non-propositional communications which we will call "emotion signals". They function both to set the whole system suddenly into a particular mode, and to maintain it tonically in that mode. We will call these "emotion modes".

Taking fear or anxiety as an example once more: anxiety may occur when a background self-preservation goal is threatened in the course of action directed towards a different goal. For instance, while watching television alone in an empty house you may hear a door open. A fright interrupts your activity. It sets the whole system into a mode of preparedness for escape or response, and maintains for a while a state of wary vigilance (cf. Gray, 1982) with various physiological accompaniments. An activity has been interrupted. As Mandler (1964) has shown, with such an interruption an emotion is likely to occur. Some theorists, e.g. Simon (1967) identify most emotion with the arousal that accompanies the interrupt signals that are necessary when managing multiple goals. Emotions, however, do not just occur with interruption. Emotionally toned moods can maintain the system in specific states, and it is a common observation that episodes of emotion can occur, and moods can persist, long after the event that elicited them is past. We suggest that the functions

of emotion modes are both to enable one priority to be exchanged for another in the system of multiple goals, and to maintain this priority until it is satisfied or abandoned.

Emotion signals provide a specific communication system which can invoke the actions of some processors and switch others off. It sets the whole system into an organised emotion mode without propositional data having to be evaluated by a high-level conscious operating system which would have to reason about an appropriate action. The emotion signal simply propagates globally through the system to set it into one of a small number of emotion modes. We imagine that these global signals may be based on non-specific neural pathways for quick responses, and perhaps on non-specific peptide or other chemical transmission in sustained responses.

Five Basic Emotions

We postulate that there is a small number of basic emotion modes which occur universally in the human species. Each has a characteristic phenomenological tone—though no meaning as such, as each is based on a non-propositional signal. On the basis of a variety of classificatory studies reviewed by Ekman, Friesen and Ellsworth (1982) one may infer that there are at least five basic emotion modes: they correspond to happiness, sadness, anxiety (or fear), anger, and disgust. One important criterion for a basic emotion is that the facial expression associated with it should be recognised panculturally. On this basis Ekman (1973) has argued that surprise may be a universal emotion, Izard (1971) that interest may be, and Panksepp (1982) that there is a specific neural circuit that subserves curiosity, and that this may be responsible for mediating surprise and interest. We would argue that surprise and interest are not proper members of the set because they are not single emotions, but may be aspects of many emotions. Surprise is elicited by a sudden unexpected event, such as the door opening in the example just given, and it can indicate an interruption and an abrupt transition into one of the basic emotion modes. Interest implies sustained attention to certain external events, and again may be a feature of other emotion modes. Empirical criteria for universal emotions, however, are only recently being agreed upon and hence the number of basic emotions may be adjusted with further evidence. Surprise/interest may be a basic emotion in the sense that we are discussing, but with properties that allow it to combine with other emotions. We concentrate on just five basic emotions here because the evidence for their universality as emotions is strongest, and because evidence on whether basic emotions can co-exist is sketchy.

There are, in addition, several non-emotional modes of the cognitive system. One such waking mode is conscious construction of a plan—a

certain kind of goal-directed thinking in which the operating system schedules lower level components. This mode, involving as it does, search, inference, and evaluation, is slow and liable to mistakes. Some of these mistakes might violate some of the system's multiple goals, including self-preservation. Sussman (1975) found he had to construct a specific and slow "careful" mode for such reasoning. A second non-emotional mode is free-association or daydreaming in which a person may be musing with memories and associations coming to mind asynchronously from a variety of sources without their being scheduled deliberately by a plan of the operating system.

We postulate that each emotional mode tends to inhibit the others. There may also be conflicts in which the system does not settle into one mode. The system of quasi-autonomous processors will require simulation on the basis of the properties of the emotion signals before its functioning can be fully specified. We imagine it will have some of the properties of parallel distributed processors (e.g. McClelland and Rumelhart, 1985).

For an emotion to occur, the cognitive system needs to be in one emotion mode or oscillating between two. The intensity of an emotion corresponds to the amount of the system entrained in a particular mode and to the consequent degree of locking into that mode.

An emotion mode is a necessary but not a sufficient condition for the full experience and expression of an emotion—the distinctive phenomenological tone, the somatic changes, the behavioural expressions, and courses of action. By itself the emotion mode based on non-propositional signals only prepares for action. In adults, the full emotion typically also includes a conscious evaluation of the juncture in planning, based on propositional signals reaching the operating system so that it is able to ascribe a meaning to the emotion mode, and so that voluntary action can be scheduled.

Less developed states also occur. For example, a person may feel a dysphoric phenomenological tone without being able to ascribe a meaning to it, or a speaker may talk loudly while leaning forward, but not until someone else asks why he or she is angry does the speaker experience the phenomenological tone.

Distinctions must be drawn between emotions and related psychological states. One distinction is with predispositions to emotion. Enduring predispositions are called temperaments.

Temporary predispositions are more difficult. The evidence of physiological and facial changes indicates that emotions are brief transitional phenomena, lasting usually for a few seconds or at most for a few minutes (Ekman, 1986). Phenomenologically, emotional states may last longer than this. So it seems best to regard the term "mood" as ambiguous. It may either refer to a temporary predisposition to emotion, or it may refer to an emotional state, perhaps of low intensity, capable of lasting for many minutes or several hours. Either of these meanings could imply a back-

ground against which more prominent, short-lasting episodes of emotion occur, and are expressed facially and physiologically.

Other distinctions are with instinctual action patterns such as a carnivore killing its prey, with motivations like aquisitiveness or hunger, and bodily states like pain. None of these states are in themselves emotional, although some motivations or bodily phenomena may lead to a transition by way of an emotional state.

We propose that a complex emotion, e.g. jealousy or remorse, is an elaboration of one of the five distinctive modes by means of the propositional meanings that are ascribed to it, and we will discuss the nature of this elaboration further.

THE JUNCTURES OF PLANS

So far we have described one thread of the argument, that emotions have evolved as a primitive means of co-ordinating a modular nervous system. We will now draw out the other main thread, extending the intuition shared by various theorists since Miller, Galanter, and Pribram (1960), Simon (1967), and Mandler (1975), that many emotions occur when planned behaviour is interrupted.

The cognitive system adopts an emotion mode at a significant juncture of a plan, i.e. typically, as Draper (1985) has pointed out, when the evaluation (conscious or unconscious) of the likely success of a plan changes. We assume that these junctures are both distinctive and recurring, so that the emotional system in mammals has evolved to recognise them and to establish distinctive responses to them. Indeed the function of these modes is to organise a transition to a new phase of planned activity directed to the priorities of the mode with associated goals and certain stored plans for dealing with what has happened. This mode is then sustained until another transition occurs.

In some cases a transition is made by default to an instinctual action, e.g. to freeze when frightened. Such default options have been wired into the system in the course of evolution as the best general plans for certain kinds of recurring juncture, where there is danger or when there is insufficient time or other resources available for reasoning carefully (cf. Sussman, 1975) in the conscious planning mode.

Table 1 indicates five distinctive types of juncture that occur generally in plans, and the transitions to new sets of goals that are typically accomplished by emotions occurring at these junctures.

The Function of Emotions in Planning

Emotions are part of the biological solution to the problem of how to plan and to carry out action aimed at satisfying multiple goals in environments which are not perfectly predictable. Examples of the multiple goals simul-

TABLE 1
Five Basic Emotions Together With Their Elicitors (the Junctures at Which They Occur) and Their Effects (the Transitions They Accomplish)

Emotion	*Juncture of Current Plan*	*State to Which Transition Occurs*
Euphoric		
Happiness	Subgoals being achieved	Continue with plan, modifying as necessary
Dysphoric		
Sadness	Failure of major plan or loss of active goal	Do nothing/search for new plan
Anxiety	Self-preservation goal threatened	Stop, attend vigilantly to environment and/or escape
Anger	Active plan frustrated	Try harder, and/or aggress
Disgust	Gustatory goal violated	Reject substance and/or withdraw

taneously pursued by mammals include: to find supplies of food and water, to hoard such supplies, to maintain oneself in proper climatic conditions, to avoid predators, to maintain territory, to find and court mating partners, to care for young, to guard one's position in the dominance hierarchy.

Because the environmental niches of sub-human mammals are somewhat unpredictable, models of the environment, although useful (see e.g. Oatley, 1974), can in principle be neither complete nor wholly accurate. Likewise, uncertainty prevents any complete dependence on predictive models in human planning. Moreover, people typically think only a step or two ahead and they respond, moment-by-moment, to the new arrangements of the environment that their actions help to create. Human plans are much more flexible than those so far explored in artificial intelligence (AI).

Most current AI programs for planning pursue a single main goal using models of the world to predict future states. They then unreel long sequences of steps ballistically, receptive only at certain moments to specific cues anticipated from the environment.

Insects seem to behave somewhat like AI programs: species-specific cues act to trigger action patterns rather in the manner of production rules, in which a simple cue elicits the performance of a specific action. The inherent dangers of relying on simple cues are absorbed by a low priority given to the survival of the individual. In mammals, the goals of individual self-preservation have higher priority. Their actions have multiple goals and new pieces of planned behaviour can be constructed according to unforeseen environmental contingencies.

Mammalian problems of scheduling behaviour are solved by devoting cognitive resources dynamically in the course of action. This process may be understood by contrasting it with the switching between fixed action patterns observed in simpler animals (Tinbergen, 1951) or AI systems based on production rules. In such systems a releaser, cue, or invoking pattern occurs, and this triggers an action pattern. In mammals, by contrast, switching may occur, not between one fixed action pattern and another, but between modes which are each associated with a small set of goal priorities. Here an emotion mode activates a set of modules that are compatible with each other. Hence an emotion mode invokes a limited suite of goals, action possibilities, and skills. This is indicated in Table 1, where each emotion calls a small range of alternative plans into consideration. This is not to say that mammals do not exhibit stereotypic behaviour, rather, that when they do, this behaviour is not an emotion.

This contrast between switching of action patterns in response to specific environmental cues, and transitions by means of emotion modes is illustrated by imagining how an insect or a car-assembly robot would act, as compared with a mammal, in an environment to which it was not adapted. An insect or robot would fail disastrously. A mammal would enter into an emotional state. Although it too might be incapacitated, it might be able to choose among some alternative actions. Likewise while one can imagine an insect or a robot in a repetitive oscillation of maladaptive behaviour, it is hard to imagine it in a conflict or a "difficulty" which seem to be the prerogative of mammals.

The small range of options invoked by an emotion mode explains some of the phenomenological quality of emotions: an emotion mode creates a sense of compulsion though with some slight flexibility. Options in an emotion mode are more narrowly focused than when in, say, the mode of free association, but action is still not completely automatic as with a fixed action pattern of an insect, or a reflex. Emotions function to focus attention on the matters of the transition: as Tomkins (1979) has argued, they amplify motivation. According to us, this occurs because an emotion mode makes some goals into figure while others become ground. Changes in the evaluation of plans' success initiate transitions into emotional states in which what has happened, and what should be done about it, remain focussed but ambiguous, and hence open to some reasoning.

Emotions: Local Difficulties and Global Problems

Why should there be both euphoric and dysphoric emotions? We propose that where ambiguity about what should be done at any given point in a plan is low, and goals of self-preservation are not threatened, mistakes and blocked paths in plans are assimilated smoothly. The system remains in the

emotion mode of "happiness". "Bugs" (mistakes in the plan) are treated as local difficulties and "patches" (new pieces of program to repair the bugs) are created for the current plan from available resources, perhaps in the way described by Sussman (1975). But when a self-preservation goal is threatened or when a plan is blocked and the problem cannot be solved using current resources, or when something happens to reveal an incompatibility among the multiple goals, the difficulty ceases to be local. The previous plan is interrupted, and there is a transition to a dysphoric emotion mode. Within this mode, ambiguities that have arisen in the evaluation of the event have to be resolved. Decisions have to be made as to whether the current plan should be abandoned altogether or only temporarily, what levels of change to it might be required, whether goals should be changed, and whether current models of the world need to be or can be revised.

Euphoric and dysphoric emotions are well illustrated by, respectively, an experienced and a novice programmer. For an experienced person, computer programming can be euphorically fascinating (see e.g. Turkle, 1984), because it is easy to attain a level of skill in which difficulties remain local. Bugs occur in a form in which the programmer can see what to do about them. Though some bugs are undeniably difficult to find, their virtue as a species is that they unambiguously invite a particular kind of solution, a patch to the existing program.

Novices, however, do not always cope with difficulties smoothly and locally. They often suffer dysphoric emotions (e.g. anxiety, anger, disgust, hopelessness). In a typical scenario the computer waits for some input while the programmer does not know what to do. Two cognitive systems are interacting—a technological system with some properties of an AI plan as discussed earlier, and a human planner used to flexibility and repair. Suchman (1985) points out that as far as the technological system is concerned nothing has gone wrong. In consequence there is no way of letting it know that anything might be wrong. The system is simply in a particular state waiting for a cue that it can recognise in that state. The human novice is distressed and confused because he or she can not use familiar means of repair, such as those used when a misunderstanding occurs in a conversation. For the human the situation is deeply ambiguous. Is one being stupid? Is one likely to be judged as such? If the wrong action is taken will that further damage the situation? Will the effort of learning the system be worth it? A potentially local difficulty has become global, and a transition takes place into a dysphoric emotional state.

THE HUMAN SOCIAL WORLD

In considering emotions in the human social world, the two threads of our argument, from co-ordination of modular systems, and from signalling the

junctures of plans, come together and we incorporate evidence from the evolution of species and from the development of the individual from childhood to adulthood.

Evolution takes place, as Lorenz (1969) has pointed out, largely where existing structures and aspects of behaviour are appropriated to new uses. In computational terms, what starts off as a side effect is developed into something functional. For example, the control structures of biochemical systems are liable to oscillate as a side effect of delays in the system. They have been pressed into service as representational oscillators capable of entraining on environmental rhythms such as the day–night cycle, and hence they provide the basis of biological rhythms (Oatley, 1978; 1985).

Emotions in adult humans seem to have resulted from lines of evolutionary and individual development in which new functions have developed for existing structures. While many writers have assumed that emotions are dependent on evolutionarily older parts of the brain, functional arguments about this evolution are rarer.

Evolution of Emotions

Evolution must have solved a set of design problems in scheduling progressively more complex nervous systems. Switching between action patterns, characteristic of insect-like invertebrates, has evolved into transitions between emotion modes characteristic of mammals. We speculate that this step was taken by an elaboration of nervous systems in which parts became progressively more specialised for specific goal-directed functions, such as specific processors to control the escape apparatus of invertebrates like squids and crayfish, activated by specific cues. Such specialisations form the basis of nervous systems composed of quasi-autonomous processors, or agencies as Minsky (1979) calls them, each with a specific goal, and in which overall organisation is hierarchical. The emotion of fear or anxiety evolved, according to this argument, in animals that had several means of escape; the choice of freezing or fleeing, the choice of fighting or fleeing.

Darwin (1872) observed that certain action patterns in lower mammals have vestigial descendants in humans even though they seem to serve no useful purpose, e.g. the sneer is a one-sided uncovering of a canine tooth. Though in humans this expression might accompany a mordant remark, it no longer prepares for a physical bite. But it may have become a communication important for the regulation of intra-specific aggression. This type of progression from innate action pattern to social communication implies an elaboration of the functions of the emotion of anger, and helps explain how the sneer has a current human function. At the same time, some expressions including those made by the mouth, which in humans is specialised for deliberate movements in speech, may have become more subject to voluntary control.

Emotions in Individual Development

Human infants, in comparison with those of other species, are born very immature. We all start life in a close relationship with another person, a caregiver, on whom we depend. We are equipped with a repertoire of expressions: crying, gurgling, and so on. Our caregivers' responses at this stage extend our immature behavioural repertoire.

To start with, as Emde argues (see e.g. Emde, 1983; Johnson, Emde, Pannabecker, Stenberg, & Davis, 1982) a baby's expressions are simply biological, perhaps reflex. But caregivers interpret them as intentional, and as emotional: the baby wants to be fed and is irritable, or is uncomfortable and sad, or is calm and happy. Moreover, adults judge photographs of infant facial expressions in the same discrete categories that are found cross-culturally: namely happiness, sadness, fear, anger, and disgust—as well as surprise and pain. Caregivers act as if such expressions are signals to make transitions appropriately into specific modes of caregiving: to feed, to comfort, to gaze lovingly. The expressions acquire significance for the child only through interactions with the caregiver. Later in childhood as Emde (1983) argues, they will be experienced as emotions by the child. The child also becomes sensitive to emotions in others and uses these as signals of "social referencing" to regulate his or her own behaviour.

Table 2 illustrates how emotions develop in the individual. The leftmost column describes the set of characteristic junctures in human plans that give rise to emotions, and the second column shows the corresponding basic emotions. In early childhood the crucial junctures typically concern the relations between caregiver and child (see Bowlby, 1969–80). The early social emotions engendered in this way are shown in the third column.

The early social emotions seem to be programmed innately to arise when crucial junctures occur, and consequently, as Ainsworth (e.g. 1967) has shown, they take similar forms in widely different cultures. As children grow older and become socialised, their planned activities diversify. They enter into mutual plans and arrangements that call for intentional commitment to others. Mutual plans are partly under the control of both participants and partly governed by conventions of their society: Certain junctures are still critical, and they still give rise to the same emotion modes. What changes as a concomitant of mutuality is that the range of cognitive interpretations of the emotion modes is extended to generate the adult social emotions. The fourth column of Table 2 gives some examples of these adult social emotions.

Moving rightwards along any line of Table 2 the reader encounters a developmental sequence of increasingly more elaborate cognitive interpretations that create emotion modes, and that become part of the actors'

TABLE 2
Examples of Social Emotions Developed on the Original Basis of Biological Emotions of a Single Actor

Junctures of Plans	*Basic Emotion*	*Infant Social Emotion*	*Adult Social Emotion*
Sub-goals being achieved	Happiness	Emotions of attachment	Sexual love, Delight
Failure of major plan or loss of active goal	Sadness	Emotions of loss	Depression, Disappointment
Self-preservation goal threatened	Anxiety	Separation anxiety	Embarrassment, Horror
Active plan frustrated	Anger	Rage	Vengefulness, Bitterness
Gustatory goal violated	Disgust	Disgust at faeces etc.	Distaste, Loathing

understanding of them. Only humans reach the last stage (in the last column) in which instinctual structures of attachment provide foundations for culture and language.

THE DEVELOPMENT OF MUTUALITY

With the exception of fear, which often occurs in modern life as the result of such events as near traffic accidents (Scherer, 1984), most emotions of interest to humans occur in the course of our relations with others. They are social emotions, and any theory of emotions must take this social dimension seriously, and not merely assume that our relations with other people are an extension of the way we treat the physical environment. We have argued that emotions are a biological solution to the problem of co-ordinating planned action with multiple goals in a world that is only partly predictable. The co-ordination of action among a social group involves a type of cognitive processing different from that used to interact with the physical world. For adults, to interact socially requires mental processes that allow the construction and execution of mutual plans, in which two cognitive systems co-operate. These processes depend crucially on each actor having a "model of the self".

There are two main questions about the nature of this model: How the nervous system might have a structure to contain such a model, and what the contents of the model might be.

As to structure, Johnson-Laird (1983a; 1983b) has argued that the human cognitive system has evolved to contain a recursively defined model

of itself. The mind is aware of itself, at least to some extent. Part of the mind's ability to construct models involves the ability to embed models of our own mind in our mind. In deliberate action, for instance, we not only act according to a goal, but we can know that we are able to do so. This ability, is re-presented in a model of the self, and indeed only with such a re-presentation would we be said to have an intention to act.

As to content, as argued by Oatley and Bolton (1985), the model of self develops from culture and language. Selman (1980) and Damon and Hart (1982) have traced the development of the sense of self. By adolescence people talk of this self as able to monitor and control some thoughts and emotions. Harris, Olthof, and Terwogt (1981) have shown that children are articulate about a range of emotions as "inner" experiences by the age of 11. We argue that only with the development of a reflective sense of self can the full set of complex emotions occur. Some of these depend on a person feeling the self to be enhanced (e.g. by falling in love), or damaged (e.g. in betrayal by others, or, as we will describe in a later example, by contradicting one's own definition of self). Thus in early childhood an individual might be anxious and clingy, but to talk of a lack of confidence depends on an adult sense of self. Though sadness is common in childhood, low self-esteem in depression is an adult experience.

It is a common assumption that representations of the self are inherently social, and first become accessible in consciousness as a result of relationships with others. Mead (1912, p. 141) wrote: "Inner consciousness is socially organized by the importation of the social organization of the outer world." Like James (e.g. 1890) he proposed that there was an aspect of the self that can be an object of thought. It corresponds to the model of self. Mead (1913, p. 145) also argued that this aspect of self acts to monitor ongoing activity, "criticising, approving and suggesting, and consciously planning." Though without the computational metaphor, Mead described some of the functions and reasons for an operating system containing a model of the whole system, including some of its goals and operations.

The content of the model of self includes an abstraction of what we have experienced in others' reactions to us. At first it is parents who hold up the social mirror. Mead (1913, pp. 146–147) states that:

> The child can think about his conduct as good or bad only as he reacts to his own acts in the remembered words of his parents . . . and the self which is a fusion of the remembered actor and this accompanying chorus is somewhat loosely organized and very clearly social. Later the inner stage changes into the forum and workshop of thought. The features and intonations of the dramatis personae fade out and the emphasis falls upon the meaning of the inner speech, the imagery becomes the barely necessary cues.

Mead thus claims, in essence, that an upbringing by adults programs new processors in the child, and one of these is the model of self, with the

consequences of being able in adulthood to talk about "looking after oneself" or of "not being able to control oneself". As Mead makes clear, the model of self contains some of the conventions of a community. It provides the means by which values are maintained and propagated from one generation to another. It is the social glue that holds a society together.

Mead (1913) went on to remark how in normal activity, self-consciousness is rare. People's actions are in register with their monitoring self, and correspond to habit, to character, to what they expect, and to what others expect of them. It is only when "an essential problem appears, there is some disintegration in this organization, and different tendencies appear in reflective thought as different voices in conflict with each other. In a sense the old self has disintegrated, and out of the moral process a new self appears (p. 147)."

This phenomenon of becoming self-conscious when a problem arises in a social plan is a typical part of the experience of adult emotion. The emotion mode generates pervasive signals which co-ordinate lower level modules and perhaps initiate bodily changes; the emotion signals focus attention and hence preoccupy the operating system. So for instance with a severe loss, a person might experience grief, and the inner dialogue is devoted to coming to terms with the loss. After an insult, a person feels angry and the inner debate may concern the means and advisability of retaliation. In general, conscious reflection arises from the critical juncture, and concerns such matters as its cause and its consequences for goals and plans. Each of these matters is usually highly ambiguous, and so the inner debate may be about whether to adjust the model of the self or the model of the other, about new plans, or new goals. Plans are evaluated in mental simulations in which conflicts are often detected among the multiple goals of the goal hierarchy. It is the propositional messages associated with conclusions from such operations that rise to consciousness, as Mead (1913) described, as voices in debate.

Mutual Plans

Many adult human plans are mutual: They are social, but unlike the attachment activities of infant and caregiver, they depend partly on conscious negotiation and cultural conventions. Mutual plans cannot be innately wired into the cognitive system; they must be created in the minds of more than one individual by implicit or explicit agreement. Such plans are among the most important that we make: in marriage, parenthood, employment, friendships etc. Many of our more intense and problematic emotions concern plans where mutuality has been sought for, set up, or assumed.

The creation of a mutual plan requires a more complex kind of operation than one to schedule actions in the physical world. One cannot model other people merely as complicated physical objects, or treat them by simple

strategies as in the theory of games. Plans become mutual when we negotiate, exchange knowledge, correct misunderstandings, and enter into shared intentions. Much of language is used in setting up, readjusting, and commenting on mutual plans and the assumed or established conventions that underlie them.

How to do Things With Promises

We have explained some of the biological mechanism underlying basic emotions, but some of the complex emotions that human beings experience require an understanding of the setting up and maintenance of mutual plans. Since the setting up of mutual plans is fundamental to social emotions, and different from planning that involves only a single actor in the physical world, we give here an example of setting up a mutual plan, and the emotions elicited by its non-completion. One way to set up a mutual plan is for one person to make a promise to another. A promise creates an obligation in the speaker and a corresponding expectation in the recipient of the promise.

Consider Searle's (1969) analysis of promising. We follow Power (1984) in naming the two actors Xavier (X), the promiser, and Yolande (Y), the promisee. Searle argues that an utterance (U) of X is a validly performed promise to perform the action (A) if, and only if, an extensive set of extra-linguistic conditions are met. These include what Searle calls the "essential" conditions for promising, as follows.

(a) X intends U to place him under an obligation to do A.
(b) X intends to produce in Y the knowledge that the utterance of U is to count as placing X under the obligation to do A.

Now imagine that Xavier has promised Yolande to call at her house to feed her cat while she departs for a week's holiday. Yolande is happy to have placed her cat in safe hands, and happy to have established a mutual plan with her friend. Xavier is working on a paper for a conference and forgets to feed the cat. A day before Yolande's return he comes across a reference to T. S. Eliot. He thinks of Eliot's poems about cats, and suddenly with a start, and a pounding of the heart, remembers his promise. He interrupts what he is doing, finds the key she has left, and rushes to her house. The cat is nowhere to be seen. The cat does not turn up. The neighbours have not seen it for days. Xavier finds himself preoccupied with thinking what to say to Yolande. He sadly contemplates the other implications of his omission, and experiences the emotion of remorse. When he finally meets Yolande his apology is accompanied by bodily disturbances and by gestures of agitated deference.

Xavier's emotion occurs partly because there has been a mismatch between his goal, the obligation to do A, and his actual behaviour, non-A.

One condition for the emotion of remorse is a sincere promise which has been broken. A promise, in Searle's sense, involves the act (A) being anticipated. In remorse non-A is remembered. Xavier would not feel remorse for the loss of Yolande's cat if he had not made any promise, if his promise had been insincere, or if he had amnesia for his omission.

To experience remorse, however, more is needed than the mismatch between the obligation or goal of doing A and the non-performance of A: it is a higher order cognitive appraisal with at least some conscious components, and based on the model of the self that played an essential part in mediating the mutually agreed plan. To experience remorse Xavier must infer that Yolande will regard his broken promise as an instance of untrustworthiness, and this perception of him by Yolande becomes discrepant with Xavier's own model of himself, that he is trustworthy. It is Xavier's understanding of Yolande's perception of him which is discrepant from his model of himself as a trustworthy person: the sort of person who would kindly offer to look after a friend's cat while she was away. Moreover this understanding of Yolande's perception of him has sufficient weight for Xavier to suffer a loss in his conception of himself. His model of himself suffers damage and he is no longer able to experience himself as a trustworthy person, at least in his relations with Yolande.

Power (1984) has provided an analysis of mutual intention as follows. In order to have a mutual intention both X and Y must intend some particular goal (G). They must assume that the other intends G, and they must assume also that the other assumes that they intend G. Assuming what the other assumes is theoretically an infinite series, but rather than cutting it at some arbitrary point, Power proposes a recursive formulation of the concept. Each actor intends his or her action (A) to achieve the goal. The two cognitive systems then include the goals and the assumptions each has of him or herself, and of the other.

Power's formulation is based on each actor assuming that the other will act in a specific plan, i.e. on inter-reliance. A slightly different formulation would be that each actor has some knowledge or beliefs about the cognitive states of the other. This modification, while losing something from Power's formulation as a competence model of mutual intention, allows a clearer view of the performance issues; the requirement that actors convey relevant knowledge about intentions on which reliance could then be based. A revised formulation of Power's analysis would then yield the following for the relevant parts of the cognitive systems of two actors who have established a mutual plan.

	For X:	For Y:
1.	X intends G	Y intends G
2.	X knows (X intends G)	Y knows (Y intends G)
3.	X knows what Y knows	Y knows what X knows

Line 2 of this analysis is not in Power's account but is important, and part of the idea that social actors must have a model of the self. It follows from this analysis that for two actors to make a mutual promise is to set up these cognitive structures. Together they must establish the mutual goal G, each person's action A to fulfil it, and a joint knowledge of everything relevant to that goal. A promise, therefore, is the paradigm of the type of speech act by which a mutual understanding, alliance or contract is established.

Mutuality is important for the theory of emotions, partly because achieving social co-operation itself creates an emotion mode (happiness), and partly because failures to achieve or sustain it have dysphoric emotional consequences. The emotion mode arising from mutuality or its breach is communicated to oneself and other persons involved. Moreover, euphoric emotions themselves can become goals to be achieved by mutual action.

As well as Searle's conditions for promising, an actor requires a model of self to make a promise or undertake a mutual intention: line 2 was added to Power's analysis because X and Y do not just intend, they must also know that they intend the mutual goal and their part in achieving it. As well as representing goals in this way, the model of self also represents expectancies about oneself, for instance that one will not break a promise. Keeping part of a mutual agreement then constitutes a significant juncture, confirming the model of self, and potentially giving rise to a euphoric emotion. Moreover the social communication of emotions leads each actor to become aware of the other's euphoric feelings, and a euphoric mutual emotion is created. Such emotions act to cement social relations.

Failure to keep an agreement, however, produces a discrepancy between outcome and the expectation generated by the self-model, hence a dysphoric emotion, such as Xavier's remorse.

Many mutual relations are established without explicit promises or acknowledgments. They arise implicitly by precedent and custom, in families, friendships, in larger communities, and even in nations. The fact that we can feel pride, disappointment, or outrage as evaluations of what happens to other people in our social groupings indicates that cognitive structures similar to those entered into explicitly also underlie our participation in these groupings, and that these too can be analysed in the kind of way that Power has shown. For instance Rawls (1972) has shown that the principle of justice as fairness in society has a similar implicit contractual basis.

THE COMPLEX EMOTIONS

Many adult emotions are complex in the sense that they are founded on a basic, non-propositional, emotion mode, but have a propositional evaluation which is social and includes reference to the model of the self.

In the foregoing section we analysed an example of a complex emotion,

remorse. It is founded on a basic mode, with its underlying phenomenological tone of sadness, physiological accompaniments, preoccupation, behavioural expressions, and so on. Such emotions depend on an appraisal in which performance is compared with that which is expected on the basis of the model of self. In remorse, an aspect of the sense of self is lost: The model of self suffers a decrement in its attributed positive characteristics.

The example of Xavier's remorse also illustrates two further points. First, as Katz (1980) has pointed out, an emotion may be part of a sequence of emotional states in which one mode gives way to another as events unfold or new evaluations occur. Phenomenologically this can appear as emotions following one another in sequence. For Xavier, anxiety was replaced by remorse as the threat that something bad might have happened was replaced by the realisation that it had.

Secondly, a complex emotion may start by being quite inchoate: Only with substantial reasoning about the situation and its implications may the full complex emotion develop as it did for Xavier as he contemplated his unreliability. A single emotion mode can thus give rise to a considerable range of complex emotions depending on the details of propositional evaluations that have been made. This phenomenon suggests an alternative account of the psychoanalytic concept of displacement (e.g. Freud, 1901). But it also implies that the range of possible interpretations of a basic emotional state is narrower than that implied by Schachter and Singer (1962). The hypothesis that autonomic arousal can be interpreted as any emotion has been disconfirmed (see e.g. Manstead and Wagner, 1981; Reisenzein, 1983). Our theory retains the idea that the experience of an emotion can change as one mode gives way to another and as shifts in the evaluation of a single mode occur. It does not predict, as did Schachter and Singer's theory, that bodily states are completely ambiguous as to the emotion modes from which they have arisen. Instead it is consistent with Ekman, Levenson and Friesen's (1983) finding that several basic emotions are physiologically distinguishable.

Basic emotions are developed from universal biological mechanisms. Complex emotions are founded on these, but plans and their evaluations vary from culture to culture and from person to person. Thus the complex emotion of remorse that we have described might be appropriate to a Western culture. By contrast Harré, Clarke and De Carlo (1985) describe how in Aquinas' discussion of emotions "*accidie*" is accorded the lengthiest treatment. In mediaeval times this emotion occurred with a failure of religious duty. Although related to remorse in being based on sadness it had cognitively the sense of loss of intimacy with God. In a cross-cultural example, Morsbach and Tyler (1976) describe "*amae*", a Japanese emotion evidently based on happiness, but with a sweetish quality of childlike dependence that occurs between adults as lovers.

CONCLUSION

We have proposed that emotions serve two principal functions. First they are a form of internal communication that sets cognitive processors into one of a small number of characteristic modes. This system of non-propositional communication has evolved as a method of changing the relative priorities of goals, and maintaining these priorities, within a parallel system of planning. The cognitive evaluation of the situations that create such junctures contributes to the phenomenology of emotional experience. Second, emotions are a form of external communication, important in the adjustment of social relations. Complex emotions characteristic of adult life arise at junctures in social plans, which often concern mutual goals and a model of self.

The theory resembles some existing theories of emotion at several points, but the constellation of its proposals is novel. Amongst its major implications are the following. There is a small number of innately determined emotion modes that underlie all emotional experiences, facial expressions, and behaviours, based on a specific cognitive signalling system. The indefinite number of complex emotions derives not from mixtures of these primitive emotions, but on propositional evaluations that interpret specific basic emotion modes. These evaluations of junctures of plans are likely to vary from one culture to another. They cannot, however, modulate just any state of arousal to produce a specific emotion. The basic emotion modes are physiologically as well as psychologically distinct.

Manuscript received 14 June 1986
Revised manuscript received 23 September 1986

REFERENCES

Ainsworth, M. D. S. (1967). *Infancy in Uganda: Infant care and the growth of attachment*. Baltimore: Johns Hopkins Press.

Bowlby, J. (1969–1980). *Attachment and loss, Vols 1, 2 and 3*. London: Hogarth Press.

Damon, W. & Hart, D. (1982). The development of self understanding from infancy through adolescence. *Child Development, 53*, 841–864.

Darwin, C. (1872). *The expression of the emotions in man and the animals*. Reprinted, Chicago: University of Chicago Press, 1965.

Dewey, J. (1895). The theory of emotions II: The significance of emotions. *Psychological Review, 2*, 13–32.

Draper, S. W. (1985). A micro-survey of ten theories of emotion. Unpublished technical memorandum, University of Sussex.

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review*. New York: Academic Press.

Ekman, P. (1986). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion*. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behaviour. In P. Ekman (Ed.), *Emotion in the human face*, (2nd ed.) Cambridge: Cambridge University Press.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous activity distinguishes among emotions. *Science*, *221*, 1208–1210.

Emde, R. N. (1983). The prerepresentational self and its affective core. *The Psychoanalytic Study of the Child*, *38*, 165–192.

Freud, S. (1901). *The psychopathology of everyday life*. English translation: Standard edition of the complete psychological works of Sigmund Freud, Vol 6, London: Hogarth Press, 1960.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Oxford: Oxford University Press.

Harré, R., Clarke, D., & De Carlo, N. (1985). *Motives and mechanisms: An introduction to the psychology of action*. London: Methuen.

Harris, P., Olthof, T., & Terwogt, M. M. (1981). Children's knowledge of emotion. *Journal of Child Psychology and Psychiatry*, *22*, 247–261.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton Century Crofts.

James, W. (1890). *The principles of psychology*. New York: Holt.

Johnson, W. F., Emde, R. N., Pannabecker, B. J., Stenberg, C. R., & Davis, M. H. (1982). Maternal perception of infant emotion from birth through 18 months. *Infant Behavior and Development*, *5*, 313–322.

Johnson-Laird, P. N. (1983a). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1983b). A computational analysis of consciousness. *Cognition and Brain Theory*, *6*, 499–508.

Katz, J. M. (1980). Discrepancy, arousal and labelling: Towards a psychosocial theory of emotion. *Sociological Inquiry*, *50*, 147–156.

Lorenz, K. (1969). Innate bases of learning. In K. H. Pribram (Ed.), *On the biology of learning*. New York: Harcourt, Brace & World.

Mandler, G. (1964). The interruption of behavior. In D. Levine (Ed.), *Nebraska symposium on motivation*, Vol. XII. Lincoln: University of Nebraska Press. pp. 163–219.

Mandler, G. (1975). *Mind and emotion*. New York: Wiley.

Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York: Norton.

Manstead, A. S. R. & Wanger, H. L. (1981). Arousal, cognition and emotion: An appraisal of two-factor theory. *Current Psychological Reviews*, *1*, 35–54.

McClelland, J. L. & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

Mead, G. H. (1912). The mechanism of social consciousness. Reprinted in A. J. Reck (Ed.), *Selected writings of George Herbert Mead*. Indianapolis: Bobbs-Merrill. 1964.

Mead, G. H. (1913). The social self. Reprinted in A. J. Reck (Ed.), *Selected writings of George Herbert Mead*. Indianapolis: Bobbs-Merrill. 1964.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.

Minsky, M. (1968). Matter, mind and models. In M. Minsky (Ed.), *Semantic information processing*, Cambridge, Mass: MIT Press.

Minsky, M. (1979). The society theory of thinking. In P. H. Winston & R. H. Brown (Eds.), Artificial Intelligence: An MIT perspective, Vol. 1. Cambridge, Mass: MIT Press.

Morsbach, H. & Tyler, W. J. (1976). Some Japanese-Western linguistic differences concern-

ing dependency needs: The case of "Amae". In R. Harré (Ed.), *Life sentences: Aspects of the social role of language*. Chichester: Wiley.

Oatley, K. (1974). Circadian rhythms and representations of the environment in motivational systems. In D. J. McFarland, (Ed.), *Motivational control systems analysis*. London: Academic Press.

Oatley, K. (1978). *Perceptions and representations: The theoretical bases of brain research*. London: Methuen.

Oatley, K. (1985). Representations of the physical and social world. In D. A. Oakley (Ed.), *Brain and mind*. London: Methuen.

Oatley, K. & Bolton, W. (1985). A social-cognitive theory of depression in reaction to life events. *Psychological Review*, *92*, 372–388.

Panksepp, J. (1982). Towards a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, *5*, 407–468.

Power, R. (1984). Mutual intention. *Journal for the Theory of Social Behaviour*, *14*, 85–102.

Rawls, J. (1972). *A theory of justice*. Oxford: Oxford University Press.

Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin*, *94*, 239–264.

Scherer, K. R. (1984). Emotion as a multi-component process: A model and some cross-cultural data. In P. Shaver (Ed.), *Review of personality and social psychology, 5: Emotions, relationships, and health*. Beverly Hills: Sage.

Schachter, S. & Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Selman, R. (1980). *The growth of interpersonal understanding*. New York: Academic Press.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, *74*, 29–39.

Sloman, A. (In press). Motives, mechanisms and emotions. *Cognition and Emotion*.

Sloman, A. & Croucher, M. (1981). Why robots will have emotions. *International joint conference on artificial intelligence*, *7*.

Suchman, L. A. (1985). *Plans and situated actions: The problem of human-machine communication*. Xerox PARC, ISL 6.

Sussman, G. J. (1975). *A computer model of skill acquisition*. New York: American Elsevier.

Tinbergen, N. (1951). *The study of instinct*. Oxford: Oxford University Press.

Tomkins, S. S. (1979). Script theory: Differential magnification of affect. *Nebraska Symposium on Motivation*, *26*, 201–240.

Turkle, S. (1984). *The second self: Computers and the human spirit*. New York: Simon & Schuster.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175.