SKETCH FOR A COGNITIVE THEORY OF THE EMOTIONS

Keith Oatley,
Experimental Psychology,
University of Sussex

and

P.N.Johnson-Laird,
MRC Applied Psychology Unit,
Cambridge

Presented to Summer Institute on Cognition-Emotion Interrelations in Colorado, 18-23 August 1985.

Cognitive Science Research Paper

Serial No. CSRP. 045

The University of Sussex,
Cognitive Studies Programme,
School of Social Sciences,
Falmer,
Brighton BN1 9QN

SKETCH FOR A COGNITIVE THEORY OF THE EMOTIONS

Keith Oatley
Experimental Psychology
University of Sussex

and

P.N. Johnson-Laird
MRC Applied Psychology Unit
Cambridge

University of Sussex Cognitive Science Research Paper Number 50. Presented to Summer Institute on Cognition-Emotion Interrelations in Colorado, 18-23 August, 1985. We very gratefully acknowledge helpful discussions, especially on the relation of emotions to human planning and the ambiguities of mistakes in plans, with Steven Draper who also kindly commented on a draft of the paper.

Abstract

A sketch is drawn of a theory that emotions are cognitively based states which coordinate quasi-autonomous processes in the nervous system. They are a biological solution of certain problems of transition between plans in systems with multiple goals. Their function is to communicate significant junctures of plans to ourselves and others in order to accomplish and maintain these transitions. There are five basic emotion modes. Complex emotions are derived from these and typically involve evaluations of junctures of social plans.

---

1. Introduction

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

To argue in this way is to say that emotions reflect aspects of the organism's evolution or development that have not been optimised by natural selection or learning. The idea that emotions might have no proper function in rational beings has continued in the thinking of many later writers. But because emotions are so ubiquitous in mammals, this idea seems implausible. In this paper we explore the alternative idea that emotions do have a rational cognitive function.

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

Our theory starts with the proposal that emotions function as a specific communications both within the individual's cognitive system and to other individuals nearby. We bring together and develop two threads of theorising: one is an argument from considering problems of communication within modular nervous systems, the other is that the matters that emotions communicate are the significant junctures of plans with multiple goals.

Within cognitive science, the dominant idea about emotions has followed Darwin, though the descriptive terms have changed: emotions are viewed as side effects of several kinds of cognitive processing. They are disturbances which accompany interrupts and discrepancies among the multiple goals and representations of cognitive sytems (Simon, 1967; Sloman and Croucher, 1981; Croucher, 1985). Our proposal, though influenced by Simon, is different. It is that emotions arise from a specific separable system evolved by active animal species as part of the means whereby an individual's multiple goals can be coordinated. Moreover, emotions are further evolved in social species such as ourselves to signal junctures in mutual plans among the members of social groups. The human skin is, as it were, permeable to emotions: an emotion like anxiety not only propagates 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.

2. 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, b). In such a system a problem arises 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 waiting for an input from the other. One solution is to build hierarchies in which processors at higher levels invoke processors at lower ones, in the way described by Miller, Galanter and Pribram (1960). This method is familar both from neurological theorising and in computer programs with many levels of embedding in which one procedure can call others as subroutines, and they in turn can call further procedures.

At the top level of the hierarchy of modules in the human cognitive system Johnson-Laird argues for 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 about emotions is that they are one of two specific kinds of communication in such systems. One kind is propositional. It makes use of signals with a propositional

content* These signals correspond to calling patterns and procedure names passing down the hierarchy to invoke lower level functions, to representations of different aspects of the world, to results and arguments of functions, and to the messages that can construct new procedures. The other kind of communication is non-propositional. It is simpler, cruder, and evolutionarily older. It propagates globally among the processors to set them into specific modes at particular Junctures of multi-goal planning sequences. These non-propositional messages are what we will call 'emotion signals'. They function both to set the whole system suddenly into a particular mode, and then to maintain it tonically in that mode. We will call these 'emotion modes'.

Taking anxiety as an example once more: anxiety may occur when a background self-preservation goal is violated in the course of action directed to a different goal. For instance, while watching television alone in an empty house you may hear a door open. This 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. Other theorists, e.g. Simon (1967) and Mandler (1984) have identified emotions as interrupt signals. We do not accept that emotions are identical to interrupts, since emotions can also maintain the system in specific states, and it is a common observation that emotions can persist long after the event that elicted them is past. We argue that the specific functions of emotions 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.

Emotion signals provide a specific communication system in the modular structure of the cognitive hierarchy. This system 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 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.

five frasic emotions. We postulate that there is a small number of basic emotion modes. Each has a characteristic phenomenological tone - though no meaning as such, since each is based on a non-propositional signal* On the basis of a variety of classificatory studies reviewed by Ekman, Friesen and Ellsworth (1982) there are Just five basic emotion modes for which there is evidence of universality: they correspond to happiness, sadness, anxiety (or fear), anger and disgust. We differ from Ekman et al's assumption that surprise and interest are proper members of the set. Surprise and Interest 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 abrupt transition into one of the basic emotion modes. Interest implies sustained attention to certain external events. (We will offer further evidence for there being Just five basic emotion modes later).

There are, in addition, a number of non-emotional modes of the cognitive system. One such waking mode is concentration on

the construction of a plan - a certain kind of goal-directed thinking in which the operating system schedules lower level components. This conscious planning 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 goals. 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, however, 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 specified. We imagine it will have some of the properties of parallel distributed systems (see e.g. McClelland and Rummelhart, 1985).

For an emotion to occur, the cognitive system needs to be in one emotion mode or perhaps oscillating between two. The intensity of an emotion may correspond 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. The full emotion depends on an evaluation of the Juncture in planning and is 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. *In* ordinary language these predispositions are called moods and temperaments depending on whether they are seen as temporary or enduring. Other distinctions are with instinctual action patterns such as a carnivore killing its prey$_f$ with motivations like aquisitiveness or hunger, and bodily states like pain. None of these states is in itself emotional, although some motivations or bodily phenomena may lead to a transitions by way of an emotional state.

We propose that a complex emotion, e.g. Jealousy or remorse, is elaborated on the basis of one of the five distinctive modes by means of the propositional meanings that are ascribed to it, (cf. also Mandler, 1984), and we will discuss this further below.

3. The junctures of plans

So far we have described one thread of the argument, that emotions have evolved as a primitive means of coordinating 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) that many emotions occur when plans are interrupted.

The cognitive system adopts one of the five distintive modes at significant junctures of plans. Evidently these junctures are both distinctive and recurring, so that the emotional system in mammals has been able 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 its associated goals, and to maintain it in that phase 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 insufficient time and resources are likely to be available for careful reasoning about solutions in the conscious planning mode.

Table 1 indicates the kinds of junctures that occur in plans and the transitions that are typically accomplished.

Table 1. The five basic emotions together with the junctures at which they occur and 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	Background self-preservation goal violated	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

It is principally the consideration of the set of junctures in plans and the range of responses to them that prompts our hypothesis that five emotion modes constitute the full set.

The function of emotions in planning. Emotions are part of the biological solution to the problem of how to plan and carry out action aimed at satisfying multiple goals in environments which are not perfectly predictable. In mammals examples of the multiple goals simultaneously pursued by individuals include the following: 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 mammals are somewhat unpredictable, models of the environment, although useful (see e.g. Oatley, 1974), can in principle be neither complete nor wholly accurate.

As to human planning, Suchman (1985) has argued that similar uncertainty rules out complete dependence on predictive models. Moreover she has observed that people typically think only a step or two ahead and 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 yet explored in artificial intelligence (AI).

Most current AI programs for planning pursue a single main goal, minimising side effects and 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 signals act to trigger action patterns rather in the manner of production rules or similar AI plans. The inherent dangers of simple calling patterns are absorbed by a low priority given to the survival of the individual. In mammals, the goals of individual self-preservation have higher priority.

For mammalian action with multiple goals and continuous modification of plans the problem of scheduling behaviour is less one of switching between action sequences than of devoting cognitive resources according to priorities of goals that might have become relevant because of unforeseen eventualities. This means that unlike the switching between fixed action sequences observed in simpler animals (e.g. Tinbergen, 1951) transitions made by mammals are between modes with associated goal priorities. What occurs is that a set of goals is activated and a suite of instinctual action patterns and perhaps skills is called into readiness. Thus in Table 1, rather than action sequences like those of AI plans being invoked a small range of alternatives is called into consideration.

The contrast between a planner which initiates action sequences according to specified environmental cues, and a system

based on transitions by means of emotion modes is illustrated by imagining how an insect or a car-assembly robot would act maladaptivly in any environment for which it was not developed, A mammal would enter into an emotional state, and perhaps be in the position to choose among some alternative actions. Moreover, where it is easily possible to imagine a pathology such as an oscillation in an insect or a robot, it is hard to imagine it in a conflict or a 'difficult' situation.

The narrow 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 focussed 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 as a kind of attentional focus onto the matters of the transition. As Tomkins (1978) has argued the effect of emotions is to amplify motivation. According to us this occurs because an emotion mode makes some goals into figure while others become ground. Cues arising at significant Junctures of plans 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 situated reasoning.

Emotions, local difficulties and global problems. Why should there be both euphoric and dysphoric emotions? We propose that where ambiguity 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[1]. There is no ambiguity about the kind of action to be taken. 'Bugs' (mistakes in the plan) are treated as local difficulties and 'patches[1] (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 violated, or when a plan is blocked and the problem cannot be solved by a patch to the current plan, or when something happens to reveal a contradiction among the multiple goals the difficulty ceases to be local. There is a transition to a dysphoric emotion mode, and the previous plan is interrupted. Often ambiguities have arisen and have to be evaluated: 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.

When unpredicted events are absorbed smoothly, by constructing patches to current plans then the system stays in either a neutral mode of conscious planning, or in a euphoric action mode with the achivement of subgoals being signalled. If a patch to the current plan will not do, the Juncture in the current plan signals a transition to a dysphoric emotion mode.

Euphoric and dysphoric emotions are well illustrated by, respectively, an experienced and a novice programmer. For an experienced person, computer programming can become 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 that 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 and 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 above, 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. No errors have occurred. 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, as for instance when 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? If the effort of learning the system is invested will it be worth it? A potenially local difficulty has become global, and a transition takes place into a dysphoric emotional state.

4. 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 also invite evidence from evolution of species and from the development of the individual from childhood to adulthood.

Evolution can only take place, as Lorenz (1969) has pointed out, 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 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 activiated 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 anxiety evolved, according to this argument, with animals that had several means of escape, or the choice of freezing or fleeing, or the choice of fighting or fleeing.

Moreover, 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 might accompany a mordant remark it no longer prepares for a physical bite. This distinctive behavioural expression has become a species-specific communication important for the regulation of intra-specific aggression. This implies a social elaboration of the functions of the emotion of anger, and helps explain how the sneer has a current human function.

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 caretaker, on whom we depend. We are equipped with a repertoire of expressions - to cry, to gurgle, and so on. Our caretakers' responses at this stage extend our immature behavioural repertoire.

To start with, as Emde (see e.g. Johnson, Emde, Pannabecker, Stenberg, & Davis, 1982) argues a baby's first expressions are reflex. But caretakers 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. Caretakers also act as if these are signals to make transitions appropriately into specific modes of caretaking, to feed, to comfort, to gaze lovingly. The expressions acquire significance for the child only through interactions with the caretaker. Later in childhood as Emde argues, they will be experienced as emotions by the child.

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 caretaker and child (see Bowlby, 1969-80). The early social emotions enegendered in this way are shown in the third column.

The early social emotions seem to arise innately 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. In mutual plans 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.

Table 2. Examples of social emotions developed on the original basis of biological emotions of a single actor

Junctures of plans	Basic emotions	Early social emotions	Adult social emotions
Sub-goal attainment	Happiness	Emotions of attachment	Sexual love, Delight
Goal deprivation	Sadness	Emotions of loss	Depression, Disappointment
Self protection goal violated	Anxiety	Separation anxiety	Embarrassment, horror
Plan frustration	Anger	Rage	Vengefulness, Bitterness
Gustatory goal violated	Disgust	Disgust at faeces etc.	Distaste, Loathing

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 which become part of the actors[1] 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. In this developmental progression we reinterpret Darwin's (1872) description of adult emotions as vestiges of childhood habit.

5. The development Qf piutuality

We have argued that emotions are a biological solution to the problem oF coordinating planned action with multiple goals in a world that is only partly predictable. The coordination of action among a social group involves a order of unpredictability beyond that of the physical world. For adults to interact socially mental processes must exist which allow the construction and execution of mutual plans. In Western society these processes depend crucially on each actor having a 'model of the self*.

This model of self develops from foundations of culture and language. Selman (1980) and Damon and Hart (1982) have shown that only as a child approaches adolescence does a sense of a 'self[1] able to monitor and control some thoughts and emotions develop fully. Harris, Olthof and Terwoot (1981) have shown that children are articulate about a range of emotions as inner experiences by

the age of 11. We argue, however, that only after adolescence and the development of a reflective sense of self, can the full set of complexly derived 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 an example below, by contradicting one's own definition of self). In childhood an individual might be anxious and clingy, but only in adulthood would that person suffer embarrassment or talk of a lack of confidence. 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 first becomes accessible in consciousness as a result of relationships with others. Mead (e.g. 1912) wrote: 'Inner consciousness is socially organized by the importation of the social organization of the outer world' (p. 141). Like James (e.g. 1892) he proposed that there was an aspect of the self that can be an object of thought. It corresponds to the model of self that we have discussed as an important component of cognitive architecture. Mead also argued that this aspect of self acts to monitor ongoing activity, 'criticising, approving and suggesting, and consciously planning' (1913, p. 145). Though bereft of 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 self is an abstraction of what we have experienced in others' reactions to us, and is critically dependent on language. At first it is parents who hold up the social mirror. '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 wrrkshop 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 1913, p. 146-147).

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 consequences implying the modular state of mentality; 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 embodies the conventions of a community, and provides the agency by which values are maintained and propagated from one generation to another: 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 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 in $grief_f$ 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. The inner debate may be about whether to adjust the model of the self or 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 such reflections which rise to consciousness, as Mead described, as voices in debate.

Mutual plana. Many human plans are mutual: they are social, but unlike the attachment activities of infant and caretaker, they depend in part on more or less conscicous negotiation. 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. People can not be modelled merely as complicated physical objects. Nor can they be treated by a simple strategy as in a game playing computer program. Our 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 upon mutual plans and the assumed or established conventions that underlie them.

How to do things with promisee. In order to understand some of the emotions that human beings experience we must understand the setting up and maintenance of such mutual plans. One way to set up a mutual plan is for one person to make a promise to another. A promise creates an obligation on the part of the speaker and a corresponding expectation in the recipient of the promise.

Consider Searle's (1969) analysis of promising. The numbers 1 to 9 in parentheses below are the numbers of $Searle^f$s conditions for correctly accomplishing a promise (pp 57-61). We paraphrase Searle slightly, and 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 if and only if the following conditions are met.

Normal Conditions:	Normal conditions must obtain, e.g. U is not a joke, Y is not deaf (1)
Propositional content:	X expresses proposition P as part of U (2) P predicates future planned act A of X (3)
Preparatory:	Y would prefer A to non-A and X believes Y would prefer A to non-A (4) It is not obvious to X or Y that X would do A in the ordinary course of events (5)
Sincerity:	X intends to do A (6)
Essential:	X intends U to place him under an obligation to do A (7) 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 (8)
Semantic:	The semantic rules of the dialect spoken by X and Y are such that U is correctly uttered only if conditions 1-8 obtain (9)

Now imagine that Xavier has promised Yolande to call at her house to feed her cat while she is away on a week's holiday. 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 was 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 he will 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. It follows from Searle's analysis that a condition for the emotion of remorse is a sincere promise broken. In a promise the act (A) is anticipated, in remorse non-A is remembered. Xavier would not feel remorse if his promise had been insincere, if he had not made any promise, or if he had amnesia for his omission.

To experience remorse, however, more is needed than the mismatch between the goal of fulfilling an obligation to do A and the non-performance of A: it is a higher order cognitive appraisal based on a model of self, which in turn has played an essential part in mediating a 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 know that the other intends G, and they must know also that the other knows that they intend G. Knowing what the other knows 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 the goal (G). Take this to include intending his or her action A to achieve G. The two cognitive systems then consist of the goals and the knowledge each has of him or herself, and of the other. The relevant parts of the cognitive systems of the two actors can be depicted as follows.

	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.

Searle's categories have been criticised as arbitary. Power's analysis suggests a principalled categorisation of the conditions for X to make a promise to Y, as follows.

Preconditions for establishing mutual goal G:	Normal conditions must obtain, e.g. U is not a joke, Y is not deaf (1) Y would prefer G to non-G and X believes Y would prefer G to non-G (4) It is not obvious to X or Y that G will occur in the ordinary course of events (5)
Committment to the intention to achieve G by means of A:	X expresses proposition P as part of U (2) U predicates future act A of X (3) X intends to do A to achieve G (6)
Conveying relevant knowledge to Y:	X intends U to place him under an obligation to achieve G by doing A (7) X intends to produce in Y the knowledge that the utterance of U is to count as placing X under the obligation to achieve G by doing A (8) The semantic rules of the dialect spoken by X and Y are such that U is correctly uttered only if conditions 1-8 obtain(9)

In a mutual promise, an analogous set of preconditions and conditions are established for Y's contribution to the plan to achieve G. A mutual promise, therefore is the paradigm of the set of speech acts by which a mutual understanding, alliance, or contract is established.

Mutuality is important for the theory of emotions, partly because achieving social cooperation 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, 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 model of self also represents expectancies, for instance a rule of not breaking a promise. Keeping part of a mutual agreement constitutes a significant juncture, confirming the model of self, and potentially giving rise to a euphoric emotion. Moreover the socially communicative attribute of emotions means that with a successfully achieved mutual goal, each actor becomes 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 and 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 (1962) has shown that the principle of justice as fairness in society has a similar implicit contractual basis.

6. The complex emotions

In the foregoing section we analysed an example of the complex emotion of 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 a cognitive appraisal in which performance is compared with that predicted on the basis of the model of self. In the case of remorse an aspect of the sense of self is lost. Many adult emotions are complex in the sense of being founded on a basic emotion mode, but having a cognitive evaluation which includes reference to the model of self.

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 evaluations occur. Phenomenologically this can appear as emotions following one another in a wavelike fashion. For Xavier anxiety was replaced by remorse as the anxiety that something bad might have happened with his omission 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 with different definitions, depending on the details of the cognitive 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 implication of their 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 the kinds of 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's discussion of emotions 'accidie' is accorded the lengthiest treatment. In medaeval times this emotion occured 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. Harre et al also describe 'amae', a Japanese emotion evidently based on happiness, but with a sweetish quality of childlike dependence and occuring between adults as lovers.

7. Conclusion

We have outlined a theory that emotions are functional and that they communicate junctures of flexible plans with multiple goals. Our intention has been to provide a Gestalt to meet certain cognitive design problems and evolutionary considerations as well as matching some salient empirical observations. Specific empirical observations related to the theory include interviewing people to see what emotions follow from what junctures, and much work of this kind has been completed under the aegis of life-event research (see e.g. Oatley and Bolton, 1985). Physiological evidence is needed as to how far each emotion mode excludes

others: Gray (1982) has shown that anxiety has some of the qualities we describe* Human and animal ethology has formed the basis of some of our conclusions though further work is needed here. Linguistic research is required to analyse whether different cultural theories as embodied in language do indicate a finite set of biological states elaborated cognitvely by appraisals in terms of junctures of social plans.

References

Ainsworth, M.D.S. (1967) Infancy In tleandai infant care and the growth of attachment. Baltimore* Johns Hopkins Press.

Bowlby, J. (1969-1980) Attachment and loss. Volg l. j≥ and %• London, Hogarth Press.

Croucher, M. (1985) A computational approach to emotions. D. Phil, thesis, University of Sussex.

Damon, W. and 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 Chigaco Press, 1965.

Ekman, P.• Friesen, W.V., and Ellsworth (1982). What emotion categories or dimensions can observers Judge from facial behaviour. In P. Ekman (Ed.) Emotion ιp the human face. Second edition, Cambridge, Cambridge University Press.

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

Freud, S. (1901) The pgychopathology off everyday life. English translation: Standard edition of the complete psychological works of Sigmund Freud, Vol 6, London, Hogarth Press, I960.

Gray, J.A. (1982) The neuropgychQlogy of anyietv: ^.n enquiry into the functions off •frfe flepto-hippQcafflPal system. Oxford, Oxford University Press.

Harre, JR., Clarke, D. and De Carlo, N. (1985) Motives and. mechanisms; &n ^ntrocjuction tp the psychology of action. London, Methuen.

Harris. P., Olthof, T. and Terwoot. M.M. (1981) Children[f]s knowledge of emotion. Journal pf Child psychology and Psychiatry. 22, 247-261.

James, W, (1892) Psychology! the br»4[ef][er>] course. New York, Holt.

Johnson, W.F.• Emde, R.N.* Pannabecker, B.J. Stenberg, C.R. and 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) ftenta;! 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 find Ifrrain Theory, 6, ft99~508.

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. Prlbram (Ed.) On the biology oJL learning. New York. Harcourt, Brace and World.

Mandler, G. (1984) Mind and bofly; psychology of emotion and

stress. New York, Norton.
Manstead, A.S.R. and Wagner, H.L. (1981) Arousal, cognition and emotion: and appraisal of two-factor theory. Current Psychological Reviews. 1, 35-54.
McClelland, J.L* and Rummelhart, D.E. (1985) Distributed memory and the representation of general and specific information. Journal of Experimental Psychology: General. lift, 159-188.
Mead, G.H. (1912) The mechanism of social consciousness. Reprinted in A.J. Reck (Ed.) Selected writings of George Herbert Mfi&dU Indianapolis, Bobbs-Merrill, 1964.
Mead, G.H. (1913) The social self. Reprinted in A.J- Reck (Ed.) Selected writings of George Herbert ftead. Indianapolis, Bobbs-Merrill, 1964.
Miller, G.A., Galanter, E. and Pribram, K. (i960) Plans &xi& the structure o£ behavior. New York, Holt, Rinehart and 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 and R.H. Brown (Eds.) Artificial Intelligence: An MIT perspective. Vol 1.. Cambridge, Mass, MIT Press.
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 representations2 tfre theoretical bases offrpaj.firesearch. London, Methuen.
Oatley, K. (1985) Representations of the physical and social world. In D.A. Oakley (Ed.) Brain and mind. London, Methuen.
Oatley, K. and Bolton, W. (1985) A social-cognitive theory of depression in reaction to life events. Psychological Review, ' 92, 372-388.
Power, R. (1984) Mutual intention. Journal for itlfi Theory o£ Social Behaviour. 14, 85-102.
Rawls, J. (1972) A theory oj£ .justice. Oxford, Oxford University Press.
Reisenzein, R. (1983) The Schachter theory of emotion: two decades later. Psychological Bulletin. 94, 239-264.
Schachter, S. and Singer, J. (1962) Cognitive, social and physiological determinants of emotional state. Psychological Review. 69t 379~399»
Searle, J.R. (1969) Speech ^cts: an ^pp^y in tft^ philosophy of la.nyya.ge. Cambridge, Cambridge University Press.
Selman, %• (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.
SIoman, A. and Croucher, M. (1981) Why robots will have emotions. international Joint Conference on Artificial Intelligence. £.
Suchman, L. (1985) Plans and situated actions: the problem of human-machine communication. Xerox PARC, ISL 6.
Sussman, G.J. (1975) A computer model of skill aouiaition New York, American Elsevier.
Tinbergen, N. (195D The study oj£ instinct. Oxford, Oxford University Press.
Tomkins, S.S. (1979) Script theory: differential magnification of affect. Nebraska Symposium pn Motivation. 26, 201-240.
Turkle, S. (1984) The second self: computers and the human spirit. New York, Simon and Schuster.