Emotion and the Dynamics of Representation Use

Michael Travers
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02139
email: `mt@media.mit.edu`

Abstract

Classical theories of mental representation are full of flaws, so much so that some researchers have abandoned the notion entirely. In this paper I attempt to synthesize theories of situated action with a new view of representation that is grounded in experience called *enactive representation*. Emotion is seen to arise from the mechanisms that allow re-enacted experiences to influence action choice. A simple implemented version of this theory is presented. I also speculate on how enactive representation can shed light on the relations between planning, infant object relations, and the development of self-representation.

1. Introduction

Recent theories of intelligent action (i.e. [Agre and Chapman 1987, Brooks 1987]) have de-emphasized the role of mental representation, instead focusing on the dynamics of interaction between a creature and its world. Much intelligent behavior, it is argued, can be directly driven from the world itself, leaving no need to represent the world internally.

Despite the persuasiveness of this work, it cannot be denied that human agents are not purely reactive -- they have some sort of state, and that is at least somewhat like the classical notion of representation in that it reflects the conditions of the outside world. The problem with representationalist thought is that it presumes, explicitly or more often implicitly, that representations are logical propositions about objective states of the world. The problems with such a view are manifold: computational, pragmatic, and philosophical arguments all suggest that the objectivist view of mind is seriously flawed [Agre 1988, Lakoff 1987]. Unfortunately, almost all computational attempts to model representation have converged upon an essentially objectivist stance.

I have attempted to develop a new view of representation that is grounded directly in experience. In this view, which I call *enactive representation,* using representations to control action is a matter not of reasoning, but of partially recreating experiential states. Furthermore, this model attempts, as far as possible, to avoid centralized control of the mind, preferring instead a radically decentralized view modeled on Minsky's Society of Mind theory [Minsky 1987]. The dynamics of representation use in the absence of a central controller are not at all obvious, and lead to some surprising results.

What emerges from this picture of a nonobjectivist, decentralized, yet representational mind is the beginnings of a theory of emotion. Emotion, I will argue, arises from the mechanisms needed to make use of represented experience in the course of action. The fact that emotion is *valenced* or signed arises from the nature of action.

The first-order version of this theory (one-step lookahead with primitive emotions) is straightforward and has been implemented in a microworld. This paper describes this as well as some speculative applications of the theory to problems of self-representation and development.

2. Enactive Representation

The basic idea of enactive representation is that representations are not objective encodings of states of the world, but subjective encodings of past experience. Representations are used to anticipate and evaluate the results of possible actions. They do this by re-enacting or simulating past experiences.

In everyday routine activity, we don't need to represent the actual state of the world because it's available to us through our senses. Situated action theory [Agre and Chapman 1987] includes the idea of *deictic representation*, which combines this sort of "leaning-on-the-world" with the language of representation. It can be argued whether such constructs are really representational since they don't involve mental state;

they are more a way of talking about the relationship between the world and an actor's dynamics.

The enactive view builds on the deictic view and attempts to integrate representations of things that aren't actually available to the senses. The primary use of representation is *imaginative*, in that it is used to perceive what isn't actually at hand. An enactive representer imaginatively re-enacts a past action in order to predict the result. Action becomes not only a matter of adaptive reaction to the world (as it is in situated action systems) but of reacting to imaginings about the world's future.

Schemas

The primitive element of enactive representation is the *schema*, loosely based on Drescher's formalization of the Piagetian schema [Drescher 1991]. A schema is a structure consisting of an action, context, and result. It can be represented graphically as in figure 1.

Both context and result represent states of the world as encoded by the creature[1] and are indicated by circles in the graphic representation. The circle on the left represents the context of the action, which specifies the states of the world in which the action might be taken. When schemas are use, the context is matched against the current world state (or a simulation of a world-state) to determine if it applies. The arrow represents the action itself, and the circle on the right represents the new state that is expected to arise if the action is taken. The declarative import of a schema is that if the action is performed when the context holds, then the result will hold after the action is completed. The procedural import is say that when the context holds, the action *should* be taken, but this is complicated by the enaction process, described below.

Schemas are the raw material of the computational system. All the creature can do is encoded in schemas. There is no global process that can examine schemas and do computations over them. This is a subtle but crucial idea that distinguishes an enactive system from a planner. In a standard planner, a planning process will search over all available operations, choosing ones whose results match goals. Enactive representations do not permit this; they cannot be "read" by a global process. Instead, results must be re-enacted, which means they must be activated in a way that is similar to the original experience that produced them.

In Drescher's model, the context and result are expressed as conjunctions of Boolean literals that represent states of sensory primitives. However, for our purposes we can extend the notion of schema to include other underlying static representational mechanisms. For instance, the context and result could be implemented by a connectionist associative memory or a more structured symbolic representation system. This paper will assume that contexts and results are represented via the K-line mechanism described in Minsky's Society of Mind theory [Minsky 1987, p. 82]. A K-line records the activation state of some set of mental *agents*. Agents in Society of Mind have both procedural and representational functions. In this case, the most relevant species of agent are *micronemes* which specify features of the world [Minsky 1987, p. 211].

The schema-based theory of enactive representations may be thought of as a dynamic extension to static representational systems that allows representation of experiential and causal relationships. The existence of a schema creates a causal relationship in the head that models a causal relationship in the world. The ongoing activity of a creature can be pictured as simultaneous real-world activity and schema activation, with the schemas continually tracking activity and creating expectations as they go.

To emphasize the fact that the context part of a schema is looking at real perceptual data, the context is drawn as an eye. This notation emphasizes that while the context is driven by actual perceptual items, the result is a representation of some situation that is not actually present to the senses. Both the condition and action are deictic in the sense of being relative to their owner, but the former is computed directly from sensory input while the latter is retrieved from memory.

Learning Schemas

Schemas are (in general) learned from experience. The schema learning problem has been addressed by Drescher and will not be covered here in any detail. His system learns to coordinate and organize sensorimotor

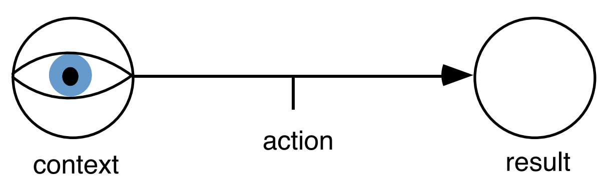

Figure 1: A basic schema

[1]A note on terminology: in this paper, I generally use the term *creature* to refer to an autonomous intelligent agent, reserving the term *agent* to refer to internal mental units, after the usage in Minsky.

information and can also derive more abstract representations of objects. Drescher's system learns through a method he calls *marginal attribution*. Every schema keeps a large table of relevance statistics, and as the schemas are activated these tables are adjusted to reflect the relevance of particular perceptual items. The disadvantages of this algorithm are the large number of values that must be stored and the requirement for a large number of interconnections between all the items.

In Drescher, the mind starts out almost blank, with no innate knowledge other than the set of actions that are available. The learning process consists in statistically determining the likely results of actions, then refining these in order to know what context conditions make the actions more reliably lead to the perceived results.

In a real creature, one might expect many more innate response mechanisms. These can be expressed by the schema formalism (although Drescher is reluctant to do so). A reflex, for instance, might consist of a schema with a context and action, but no result, since the result is not needed to produce the reflex action.

Using schemas

Schemas by themselves can specify what to do, if they are being used in a purely reactive way as situation-action rules. In this case, the result component of the schema is not used. The real power of schemas comes from being able to predict the result of hypothetical actions. In order to make use of this power, additional mechanisms are needed to allow these predictions to have an effect on action choice.

The result allows schemas to be used other than as reactive rules, in a manner which I call *imagination mode*. In imagination mode, the schema is still activated when its context matches the current situation, but instead of taking the action, the result is activated almost as if the creature were experiencing the result state. This means that the creature effectively re-experiences the results of an action without actually taking it. The idea is that imagining a future experience

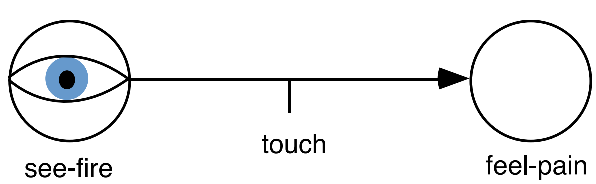

Fig 2: Schema encoding a nasty experience

(for example, receiving a painful stimulus) is much like actually undergoing the experience again.

However, we don't want the creature to believe that it has really achieved the result, so imagined futures shouldn't be completely identical to experience. One way to accomplish this might be through some form of gating that restricts the scope of the result activation. This might be managed by something like Minsky's level-bands [Minsky 1987, p. 86] so that the memory is restricted to general rather than specific qualities.

To make this recreated experience useful, it must have an effect on action. More precisely, the activation of a schema's result must somehow have an effect on whether or not the schema's action is actually taken. This effect might be positive or negative depending upon the nature of the result. A desirable result should enhance the likelihood that the schema's action will be taken, while an undesirable result should suppress it.

I hypothesize that there are control mechanisms that allow agents to influence the action-taking function of schemas. These are attached to particular agents based on their desirability. An undesirable result (say, pain) will be connected to a control mechanism that suppresses the action of the schema under consideration. A desirable result, on the other hand, will excite the schema so that the action is more likely to be taken. When schemas are in competition for activation, these mechanisms will affect their relative chances.

Note that the effect we desire is temporal rearrangement of action: whereas the original experience happened as context-action-results, we need to imagine the results *before* the action is taken.

Example: For instance, take a simple case of an animal learning that fire is dangerous. The present theory would have it that the animal's training experience (of seeing a fire, touching it, and feeling pain) is recorded as a schema (see figure 2).

Now, when the animal sees fire, the schema is activated in imagination mode, activating agents in the result portion of the schema. As a result, the animal undergoes a partial experience of pain. These agents are connected to inhibiting mechanisms that cause the action associated with the schema to be suppressed (see figure 3).

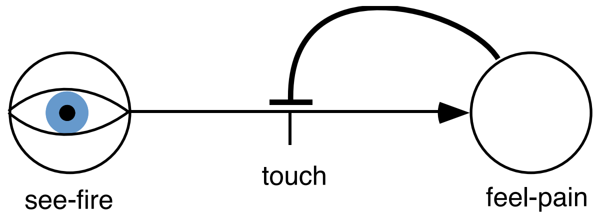

Fig 3: Schema preventing itself from acting

So, as a result of the past experience as recorded in the schema, the animal can avoid taking a potentially harmful action. If the animal has other schemas that can be activated in this situation, they will now have a better chance of taking their actions.

Summary: The behavior of the animal has been improved, but what is this like for the animal? What can we say about its experience? What enactive representation tries to capture is the experiential connection between past experience and present thought. An animal confronted with fire must, in essence, scare itself, and the easiest way to do that is to reproduce the past experience. The important point here is that remembering the experience is like partially re-experiencing it. The agents that comprise the representation of the result get partially activated. Being afraid of something is akin to being hurt by it again. The nature of representation is not symbolic or declarative, but experiential.

The activity of the mind parallels the activity of the creature. Schemas throw up projections of possible futures; which are in some sense re-creations of past experience. These projections have an effect on what action actually gets taken by means of excitatory and inhibitory mechanisms.

The distinction between this view of action and that of classical planning is that the action and thought are not radically separated. The strategy of planning is to make an internal model of the world, a model which can be freely examined and manipulated by a centralized mind in the safety of the head. Once a plan is constructed in a head, action is simply a matter of spitting it out through an execution module. (See [Agre 1988] for a detailed critique of the planning view).

Because traditional representation implies a detachment from the world, the enactive representation strategy is to look at the ways in which thought and action are integrated. Theories of intelligent action should be composed of minimal mechanisms that can learn and make use of real-world experience, rather than airtight formalisms.

Emotion

The term emotion is perhaps too vague. It isn't clear what a computational system that claims to be a theory of emotion should actually accomplish. Nevertheless, it appears that no other term is better at denoting reactions that are conscious but not voluntary, usually but not necessarily expressed behaviorally, and having a positive or negative valence, that is, involving an evaluation of some condition or entity as desireable or not [Ortony, et al. 1988].

The mechanisms of the previous section form a basis for a theory of emotion. Emotions are dynamic responses to both situations and to representations of past situations. The two simplest emotions are fear and desire, based on negative and positive reactions to schema representations. A schema that projects a negative future will suppress the corresponding action (as in the example above), and this dynamic of being reminded of a bad experience and inhibiting action we will interpret as fear. On the other hand, a schema that leads to a positive result will try to take its action, which can be interpreted as positive attraction or desire. The valenced nature of emotion arises from the two types of action control mechanism: action-enhancing or action-suppressing.

These basic emotions are neither mysterious nor particularly hard to explain, being functionally grounded in the adaptational necessity. Why then, do we think of emotions as complex and hard to describe? Emotions get more complicated when the reactions are to representations of the self or of other people. It is this capability in conjunction with the simple emotional reactions that produces the complex responses of human emotion. For instance, guilt can be thought of as a negative reaction to the possibility of some other person evaluating one's self negatively. This characterization of a single emotion is simplistic, to be sure, but this type of complex representational dynamic matches the complexity and dynamics of emotion.

The interaction of imagination mode and the inhibitory mechanism can perhaps explain Freudian repression. The inhibition of a schema will, presumably, also inhibit the imaginary activation of the result. In other words, as soon as something bad is thought of, not only is the action leading to it inhibited, but the very thought itself is extinguished. Schemas that lead to extremely unpleasant results then inhibit themselves so thoroughly that the imagined result is almost inaccessible, because it turns itself off as soon as it is remembered. Yet the inhibition of action is certainly a noticeable effect. These dynamics suggest that the Freudian unconscious may be constructed out of self-inhibiting schemas. Making this actually work presumes that there is some hysteresis to inhibition, or else the creature would immediately forget to be scared!

Conflict and choice

Rational action is usually conceived of as a matter of making choices: given a set of actions, perform the one that leads to the most desirable outcome. The mechanisms presented thus far only work with single schemas. If more than one schema applies to a situation, how does a single action get chosen?

In general, there needs to be some arbitration mechanism that mediates between actions that cannot be performed in parallel. Because motor effectors are used for more than one behavior, the outputs of behavior modules must converge somewhere, and some mechanism must control which behavior has control of the effector. One solution is to have the activation value of the behaviors take on a range of continuous values and use some form of winner-take-all competition among them, based on the desirability of the results of the schemas involved.

The problem is that if schemas activate their results in parallel in imagination mode, things can get confused. The control mechanisms have no way of knowing which schema led to the activation of a particular result-agent, and should thus be enhanced or suppressed. One solution is to activate the schemas serially. This requires significantly more control machinery than we have yet had to postulate. Not only must there be some central controller to activate the schemas in parallel, but each schema must somehow remember its level of activation so that it may be compared after each schema has had a chance to imagine its result. This is unfortunate, given the goal of avoiding centralization.

One way of avoiding this sort of choice is suggested by the phenomenon of displacement behavior from ethology, as well as a more complex version of the idea from *Society of Mind*. The general tactic is to exploit conflicts between possible actions as opportunities for learning. In other words, when behaviors conflict, rather than making a choice between them, defer the choice to some other mechanism.

Animals can often exhibit odd behavior in situations where they are faced with making a choice between two conflicting but equally desirable actions. In such a situation an animal will often do *neither* of the contending actions, but instead displays a third, seemingly arbitrary behavior. One classic example is birds that confront an aggressor on the border of their territory [Tinbergen 1951]. Rather than fight or flee, they will occasionally display grooming or nest-building behavior, although none of the usual triggers for those behaviors are present.

Apparently the two conflicting behaviors, being incompatible, are suppressing each other, which allows the third, unrelated behavior to emerge. It is not clear if this dynamic, by itself, is adaptational in an evolutionary sense. However, it appears as if these displacement displays, once established, are seized upon by evolution and start to form the basis of social communication systems. That is, they become *ritualized*. Their very disutility allows them to be adapted for other purposes.

The Principle of Non-compromise [Minsky 1987, p. 33] is a more sophisticated way of dealing with conflicts. This theory depends on a hierarchically ordered network of actions. Conflict between actions on one level, so the principle states, will result in the subtree as a whole receiving less priority, presumably because it is no longer effective at achieving its goal.

In displacement, two conflicting behaviors at the same level of a hierarchy inhibit each other, allowing a third behavior, also at the same level, to emerge. In a non-compromise situation, the third behavior is at a higher level, subsuming the first two behaviors.

The implications of this for enactive representation systems are that it may be possible to avoid search.

Chaining

First-order enactive representation only lets the creature look one step ahead. In order to use these schemas to control more complex activity, additional mechanisms will be required to keep track of intermediate states.

Looking more than one step ahead requires chaining schemas—the imagined results of one have to match the context of another. The second schema cannot look at the real world, as usual, but must look at an image of the world as called up by the initial schema (see figure 4). Chaining requires simulation, in some sense, of the intermediate state ([Waltz 1990]also notes this).

Keeping track of multiple stages in an orderly manner is problematic, especially when choices are involved—how do we keep track of the intermediate states and distinguish steps? Classical planners use stacks or equivalent data structures. Enactive representers don't have those, at least, not as low-level hardware. How do we keep track of what's going on? No clear answer is evident. It may be that in actual activity, it isn't necessary to distinguish between, say, results that stem from different segments of an envisioned act. Or it may be that K-line-like mechanisms could keep track of these intermediate states, one per step. If the K-lines are actual hardware, then this would put a limit on lookahead. But it's likely that not more than 2 or 3 stages can be reliably kept track of in ordinary activity. This corresponds to our experience of normal life, where plans are not usually worked out in detail more than a couple of steps in advance. Doing so usually requires specialized cognitive tools (such as writing).

3. Summary

Emotion and Cognition are Intertwined at a Basic Level

Emotion has traditionally been a middle term in western thought, mediating between pure reason and the external world. Emotion itself has both positive and negative valuations. It's often opposed to the positive notion of reasoned thought, but also may be opposed to a negatively-valued concept of uninvolvement. To be emotional is to be unreasoning, but to be unemotional is to be cold, detached, and estranged□[Lutz 1988]. Emotion is seen both as an vaguely archaic appendage to thought, and the very seat of being. AI, of course, has tended to focus on reason, relegate emotion to the margins of theory, and ignore the issue of estrangement from the world.

The separation between reason and emotion corresponds to the separation of fact and value that is emphasized by scientific thought. This separation has been reproduced within AI research, in which reasoning is paramount, and emotions and values appear, if at all, simply as arbitrary inputs to general reasoning engines.

Rather than separating sensation, representation, action, and emotion into separate modules, the enactive view sees them as deeply intertwined. Representation only makes sense as a guide to action, and it's obvious that all forms of representation, particularly early ones, are not neutral but deeply imbued with values.

Representation is Experiential rather than Objective

The classical model for representation is first-order logic with model-theoretic semantics. Most efforts within AI fall under the sway of this paradigm, and thus take on the many assumptions of objectivism: that the world consists of objectively individuated objects, that mental tokens correspond unproblematically to these individual objects and their properties, that mental contents are primarily statements about factual states of the world, and that mental activity is primarily deducing further valid statements about the state of the world. None of these presumptions are sensible. The objectivist view is so entrenched in our thinking that alternatives are sometimes hard to visualize. The theory of enactive representation is an attempt to provide an alternative to objectivism based on the dynamics of experience and of

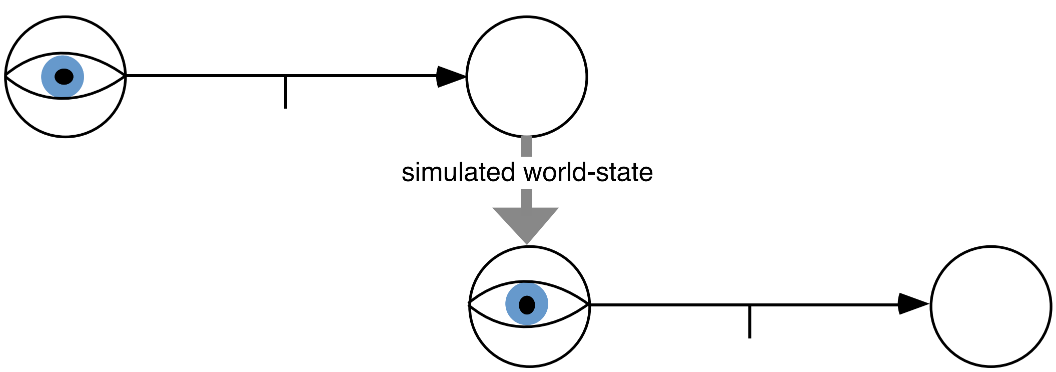

Fig 4: Two chained schemas. The second schema is no longer being triggered by actual perception, but by imagined perceptions.

representation use.

Implementation

This first-order theory of enactive representation, which is capable of looking one step ahead, has been implemented in a simple microworld. This implementation includes a simplified version of Drescher's schema learning algorithm and a world simulator similar to an earlier animal-behavior simulation system [Travers 1988]. The system is capable of learning simple schemas such as "touching a yellow object results in pain" or "eating a blue object results in pleasure" and using these schemas for enactive projection and choosing actions.

4. Extensions and Speculations

In this section I attempt to sketch out how enactive representation might be fruitfully applied to problems in development and self-representation. Children early on form emotional attachments to objects. The nature and implications of these attachments have been explored by object relations theory; we will see that there is some resonance between this theory and enactive representation. Other theorists have shown how the ability to represent the self arises from social interaction. I will examine these theories in light of enactive representation. It is my conjecture that the ability to chain schemas and thus imagine farther into the future is based on the ability to represent the self, as both require the underlying ability to imagine states-of-the-world other than those available to the senses. This cluster of abilities also makes it possible to experience more complex emotions.

Some of the language used in this section assumes that representation is in terms of prototypes, and that one object's representation can serve as the prototype for another's. Prototype-based representation is advocated by [Lakoff 1987]. It has not been explored much within AI (but see Haase, 1991 #636).

Object Relations

The object relations theorists (Melanie Klein and D. W. Winnicott being the most prominent) observed interactions between very young children and external objects (parents and playthings) and concluded that development was intimately bound up in these external relationships and their internalized representations. It should be emphasized that "objects" here refers not to neutral physical objects, but objects of emotional attachment (people or things like teddy bears), or more accurately to the relationship of a subject (another person) to these objects.

Melanie Klein [Klein 1986] held that the prototypical object of object relationship was the mother's breast. Early in infancy, children develop emotional attitudes towards the breast as the primary source of satisfaction (when it provides milk) and disappointment (when empty or denied). Love and hate are then defined as the basic emotional attitudes engendered by the object. Klein believed that this dual attitude resulted in a "splitting" of the breast (we would say the representation of the breast) into a "good" and "bad" breast. These basic reactions to the prototypical object are the roots of all other emotional and representational activity.

Other processes playing a key role in development are *introjection* and *identification*. Introjection involves taking characteristics of the object and internalizing them, building them into one's self-representation or ego, thus identifying with them.

As the infant learns to introject (or represent) the object, it becomes less dependent on the actual physical presence of the object. It can experience emotional relating, not only to the object immediately at hand, but also to its introjected representations.

Between the breast and the introjected object lie *transitional objects* [Winnicott 1971] — items such as dolls or blankets that are external (like the breast) but can be incorporated into the independent life of the child (like mental objects). According to Winnicott, children engage with such objects in relations of both love and destructiveness, essentially acting out aspects of their object relations but from a position that offers them greater autonomy than with the breast.

As the child develops and learns to make more realistic object representations, the physical identity of the good and bad object can be a source of emotional turmoil. One can't keep the good and repress the bad if both representations ultimately refer to the same physical object. Integration of actions and object-feelings can also be problematic: for instance, at some stage the child realizes that while food may be "good," the enjoyment of it results in its destruction. From the basic duality of emotional reactions (love/hate), and the need to identify with and introject objects, come the basic psychological disorders (and the "vicissitudes of the normal self" as well).

Enactive representation and object relations

Both object relations theory and enactive representation are based on the existence of valenced reactions at the earliest stages of development. Whereas enactive representation views these reactions as pragmatic necessities for getting about in the world, object relations focuses on the ambiguities that arise when reactions of both valences must co-exist.

When schemas are invoked as projections, they can invoke object representations which are the traces of past object-experiences. We can imagine an implementation of Kleinian object-relations that involves schemas that are based on early experiences with the breast and result in both good and bad results, resulting in conflicting expectations and a conflict situation which must be handled by some other mechanism (see the earlier discussion). Future object representations use the breast representation and associated schemas as their prototypes, carrying the emotional split forward into other areas of life.

Obviously this sketch of infant behavior and emotion must be incomplete and speculative. For one thing, there are seemingly innate reflexes involved, such as rooting and sucking. Perhaps these (presumably non-representational) reflexes serve to "bootstrap" representational schemas, by implementing behaviors such as exploration and interaction that are likely to lead to further learning.

The Other and the Self

The social perspective on development holds that self-modelling evolves from the ability to model others [Mead 1934]. In representational terms, this means that the representation of "the generalized other" (Mead's term) is used as a prototype for the representation of self. This underlies the process of interalization, in which parental judgements become those of the child.

Kegan's Developmental Theory: Like the present work, Kegan [Kegan 1982] is trying to build a theory that integrates affect and intellect, avoiding subjugating one concern to the other. He views the growth of personality as a successive development of different forms of object relations. This development happens in stages. People are driven from one stage to the next by "developmental crises" or "natural emergencies", periods in which the previous relationship between subject and object becomes unsatisfactory and a new one must be negotiated. Kegan regards the word "object" according to its Latin roots, which indicate that it can be interpreted as a process of throwing-out, or differentiating something from the self. His view of development is essentailly Piagetian, although it also integrates elements from psychoanalysis, notably object relations.

Each stage of development involves construction of a new subject, which takes the subject of the previous stage as its object (in Society of Mind terms this can be envisioned as growth of new layers of managerial agents, although this does not capture all of what Kegan is trying to convey). This development is *dialectical*, in that alternating stages involve either the differentiation or the integration of the self with the exterior world.

For instance, a newborn child, at Piaget's sensorimotor level, "is unable to distinguish between itself and anything else in the world." [Kegan 1982. p. 30] whereas a child at the next stage (preoperational) can differentiate itself from objects but cannot comprehend that the world will appear differently to other people. Further stages of development involve learning to make more sophisticated distinctions between subject and object and changes in the way they relate.

Chaining depends on self-representation

I conjecture that the ability to represent the self underlies the ability to keep track of intermediate states in the process of chaining. Furthermore, both derive from the social context of action and development. You learn to represent yourself by observing how others represent you.

Both chaining and self-representation involve taking a new perspective. In an enactive system, this means that imagined representations get fed back into perception, so that schemas that normally react to the perceived world can react to an imagined world.

In the case of chaining, this imagined world is the future state of the imaginer. In the case of self-representation, the imagined world is the world of an outside observer — one which includes the imaginer as an object.

In development, the tendency and ability of adults to take the child as an object is internalized so that the child can take itself as an object. In enactive terms, this means that the child must imagine herself, then allow her schemas to operate upon those imaginings. Development proceeds by successive stages as the ability to do this sort of representation grows. The functional motivation for this ability is that it underlies the ability to see and plan farther into the future through chaining.

Processes like self-representation are very difficult to talk about sensibly, since our language assumes an existing world of actors and selves. Nevertheless, introducing them at an early stage may provide a better foundation for thinking about action, representation, and emotion than the more traditional logic-based techniques. Rather than assuming an adult rational actor, we should take a developmental approach, and try to imagine how a mostly-reactive child can come to imagine the world and itself.

Complex emotions from complex representations

The first-order enactive system has but two basic emotions, which we've identified as fear and desire. More complex emotions retain the valenced nature of the basic emotions, but complicate them by applying them to more elaborated representations, especially those involving the self and others.

Emotions, then, consist of valenced reactions to complex representational systems, especially representations for *persons* — both the self and "the generalized Other".

Complex emotions can be pictured as complex dynamics involving self, others, look-ahead mechanisms, and the positive or negative valuations supplied by the simple action-selection mechanism. For instance, if a child has the ability to represent another's disapproval, it can experience an emotion something like shame. Actions that lead to this sort of disapproval evoke an inhibitor, since the disapproval is felt as unpleasant. Other possible emotional dynamics:

• Security (dependence): if I want something, the Other will satisfy it for me.

• Envy: I view the Other as possessing something desirable, whereas Self does not, and there is no possible actions that lead to a state of having.

By combining enaction with self-other representation, a schematic form for many complex emotions can be sketched out. Viewing the dynamics in conjunction with the necessity to separate self and other provides a richer view of emotion than purely typological approaches such as [Ortony, et al. 1988]. Emotions are not states, but dynamic processes.

Acknowledgement

Edith Ackermann, Amy Bruckman, Marc Davis, Pattie Maes, Kevin McGee, and Carol Strohecker read drafts and helped dispel some of the murkiness.

This work was supported in part by Apple Computer and Mitsubishi Electric.

References

Agre, P. E. *The Dynamic Structures of Everyday Life*. 1988. AI Laboratory Technical Report +++.

Agre, P. E. and D. Chapman. "Pengi: An Implementation of a Theory of Situated Action." In *Proceedings of AAAI-87*. 1987.

Brooks, R. A. "Intelligence Without Representation." (1987):

Drescher, G. L. *Made-up Minds: A Constructivist Approach to Artificial Intelligence*. Cambridge, MA: MIT Press, 1991.

Haase, K. "Making Clouds from Cement." 1991.

Kegan, R. *The Evolving Self*. Cambridge, Massachusttes: Harvard University Press, 1982.

Klein, M. *The Selected Melanie Klein*. New York: The Free Press, 1986.

Lakoff, G. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press, 1987.

Lutz, C. A. *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll & Their Challenge to Western Theory*. Chicago: University of Chicago Press, 1988.

Mead, G. H. *Mind, Self, & Society*. Vol. 1. Works of George Herbert mead, ed. C. W. Morris. Chicago: University of Chicago Press, 1934.

Minsky, M. *Society of Mind*. New York: Simon & Schuster, 1987.

Ortony, A., G. L. Clore and A. Collins. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press, 1988.

Tinbergen, N. *The Study of Instinct*. Oxford: Oxford University Press, 1951.

Travers, M. "Agar: An Animal Construction Kit." MS Thesis, MIT Media Lab, 1988.

Waltz, D. L. "Eight Principles for Building an Intelligent Robot." In *First International Conference on Simulation of Adaptive Behavior in Paris*, MIT Press, 462-464, 1990.

Winnicott, D. W. *Playing and Reality*. London: Routledge, 1971.