What AI Needs is a Theory of the Functions of Emotions

Hartvig Dahl
Department of Psychiatry; SUNY Health Science Center
450 Clarkson Avenue, Brooklyn, NY 11203 USA
dahl@hscbklyn.edu

Virginia Teller
Department of Computer Science; Hunter College CUNY
695 Park Avenue, New York, NY 10021 USA
teller@roz.hunter.cuny.edu

Abstract

AI needs a formal functional theory of emotions to represent in computer programs our knowledge of emotions. In the theory we propose, emotions are a special class of Intentional states with structural components and properties similar to those of the traditional somatic appetites of thirst, hunger, and sex. These were originally part of a hardwired, phylogenetically adapted, nonverbal feedback system for implicitly conveying information about these states both among and within individual members of the species. A classification system provides **two major functional classes** of emotions, (1) those serving as Appetitive Wishes toward objects, and (2) those serving as Beliefs about the status of fulfillment of those and other significant wishes. Thus, emotions such as **Anger** or **Fear** indicate a wish to attack or escape from some object or situation, while **Love** or **Surprise** indicate wishes to care about or explore an object or situation. Emotional wishes, like their somatic brethren, require Consummatory Acts for their fulfillment. The result of these acts are emotions such **Anxiety** or **Depression**, which indicate Beliefs that the relevant wishes will be hard or impossible to satisfy, or **Contentment** or **Elation**, which function as Beliefs that the wishes have been or are being fulfilled. Together, emotional wishes and beliefs form a comprehensive wish-belief information feedback system with manifold causal consequences. Potential applications include agent-oriented problem solving, animation, and human-computer interaction.

Introduction

AI needs a coherent theory of the functions of emotions to represent our commonsense knowledge of emotions in computer programs. As Johnson-Laird and Oatley (1992) stated, ". . . emotions should have a function that could be embodied in a system based either on carbon-like humans or on silicon-like computers" (p. 204). If a *prima facie* goal of artificial intelligence is to understand how we **know** the world around us and ourselves in it, then it is essential that AI include a theory of emotions. Here we formalize this commonsense knowledge (cf. Minsky 1975) as a rational, scientific account of emotions that are biologically wired into humans and that can, in principle, be implemented in computer programs. This is not a theory of how emotions are expressed but of the wishes and beliefs that give rise to expressions of emotion in language, actions, facial expressions and physiology.

In the absence of a theory that explains the functions of emotions, we are left with the unsatisfying result described by Dennett (1996), who claimed that if emotional experience is subtracted from consciousness, then "nothing is left beyond a weird conviction (in some people) that there is some ineffable residue of 'qualitative content' bereft of all powers to move us, delight us, annoy us, remind us of anything" (p. 6).

Emotions are Cognitive Processes

Our claim is that a vast amount of our knowledge about ourselves, our goals, the behavior of other human beings, and that all pervasive but elusive thing called common sense, is based on the *implicit* information content of emotions. An empirically testable theory of emotions, such as the one proposed here, can help specify this implicit knowledge.

The theory we outline is quintessentially *cognitive* in the sense of *perceiving* and *knowing*. It views emotions as having evolved as a (the?) primary means for our mammalian ancestors to communicate with and understand fundamentally important intentions of other members of their own species. In this sense emotions constitute a basic *information processing, i.e., cognitive,* capacity. Moreover, emotions are the first language of every human infant before symbolic language is acquired. The commonsense knowledge that each of us has of our own and others' emotions underlies all of our human interactions. A successful emotion theory must account for the *cognitive* function of emotions and explain, rather than take for granted, our knowledge of classes of emotions with radically different functions. This version of the theory[1] is based upon three basic propositions: (1) a 3-dimensional classification scheme of emotions, (2) the intentional (cognitive)

[1]For the previous versions, and background and support for some of the major claims see Dahl (1978, 1979, 1983, 1995) and Dahl and Stengel (1978).

concepts of *wishes and beliefs,*[2] and (3) the biologically rooted concept of *appetites*. The ground covered includes:

1. A *3-dimensional classification* scheme;
2. Basic definitions of *wish, pleasure, unpleasure,* and *appetite*;
3. *Two major functional classes* of emotions;
4. *A causal feedback model* of the two classes;
5. The *relationship of the model to commonsense knowledge* of emotions.

A 3-Dimensional Classification of Emotions

Any theory of emotions must include some explanation of both their similarities and their diversity; such an account implies, at a minimum, some system of classification. This system is an adaptation of the *n*-dimensional scheme that de Rivera (1962) used in his "decision" theory of emotions. The three polarities are the same as those that Freud (1915) claimed are basic to mental life: Subject-Object, Pleasure-Unpleasure, and Active-Passive. Our adaptation employs the following terms:[3]

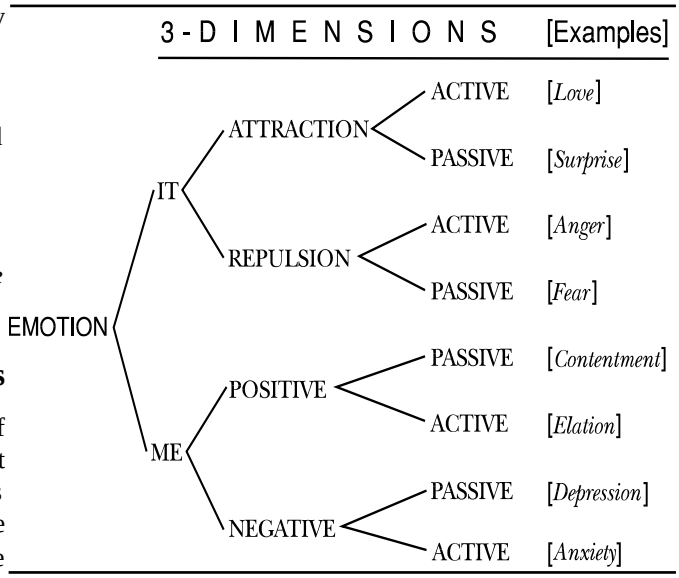

Figure 1

Orientation	[IT-ME]
Valence	[ATTRACTION/REPULSION-POSITIVE/NEGATIVE]
Control	[ACTIVE/PASSIVE]

A classificatory tree showing the results of the intersections of these dimensions is illustrated in Figure 1 together with typical examples of emotion names for each of the eight resulting categories.

If such a scheme has any validity it surely implies that ordinary people have some kind of internal representation of these dimensions and ought to be able to use them to classify emotions. In fact, in his dissertation de Rivera showed that 20 judges could reliably classify 188 emotion words based primarily on their knowledge of definitions of these abstract dimensions. Dahl and Stengel (1978) replicated and extended his empirical classification using the above three dimensions. They gave 58 judges definitions of each dimension and had them classify 400 emotion words to see if they actually did share both the implicit knowledge of the dimensions and knowledge of the internal states referred to by each emotion label. The reliabilities (coefficient alphas) of these judgments by 58 judges for the three dimensions were .95, .99+, and .97, and the intercorrelations among the dimensions were nil, providing important evidence for the empirical independence of the dimensions. Approximately 65% of the judged words were decided at $p < .05$ on all three dimensions; for 153 words the splits were significant at $p < .001$ on each of the three dimensions (a joint chance probability of <1 in 1 billion). Moreover the judges' ages and sex were uncorrelated with their choices.[4]

---

[2]These terms from our everyday common sense were first given a more formal status in Heider's (1958) "commonsense" or "naive" psychology and later in "folk" psychology (Stich 1983) and intentional system theories, e.g., Dennett (1978, 1987, 1988), Searle (1983) and many others.

[3]Among those who have proposed n-dimensional schemes only de Rivera among emotion researchers, along with Descartes (see Stone 1980) and Freud, has stressed the importance of the Subject-Object (IT-ME) dimension. No theoretical (as distinguished from commonsense) rationale for this distinction is provided by any other major theory except perhaps that of Pribram and Melges (1969). The terms "orientation," "valence," and "control" are intended to capture the essential properties of the dimensions.

[4]A complete list of the words (with the distribution of the judges' choices on each dimension) and a figure showing the classifications appeared in Dahl and Stengel (1978). See also Dahl, Hölzer & Berry (1992).

Thus, despite the obvious fact that each person has *direct* access only to his own particular emotional states *and* in principle cannot have such access to another's internal state, and, despite the fact that each person has his own unique set of memories derived from his own developmental interactions, it seems necessary to assume substantial shared experiential referents in order to account for the judges' shared variance and agreement.

However, even if we accept the tidy simplicity of this scheme, there is a certain arbitrariness to all classifications of emotions, as James (1890) noted when he suggested that "the only question would be, does this grouping or that suit our purposes best" (p. 485)? For purposes of outlining this theory, there were two results of special interest.

The first was that these abstract dimensions produced 8 main categories (as shown in Figure 1) and 4 intermediate categories (not shown) which together include most of the emotions that others have regarded as fundamental, based on a variety of criteria such as behavioral expressions (Darwin 1872), instincts (McDougall 1923), facial expressions (Tomkins 1970; Izard 1971; Ekman 1973) and other (Plutchik 1962; Davitz 1969; Ortony, Clore , and Foss 1987).

The second result was that the classification provided, as we shall show, *a principled distinction* for two major *functional* classes of emotions, here called **IT** and **ME**. Skeptics who find the *n*-dimensional classification implausible on other grounds should remind themselves that in the course of evolution just such *n*-dimensional computations were selected very early, allowing mobile animals with multiple senses to orient themselves in space and time by vision, smell and sound. It is conceivable that evolution, having once selected whatever powerful computational methods underlay these capacities, might also use *n*-dimensional computational strategies for solving new problems such as the apparent need for conspecifics both to express and to recognize each other's intentions, i.e., wishes and beliefs.

Basic Definitions and Characteristics

Wish, Pleasure and *Unpleasure*

The theory of emotions as wishes and beliefs rests firmly on a definition of a *wish* as *an attempt to achieve perceptual identity and/or symbolic equivalence with a previous experience of satisfaction*.[5] *Pleasure*, in this model, is the *satisfaction of a wish* and *unpleasure* is the *nonsatisfaction of a wish*. These definitions have several decisive implications, some obvious, some not quite so. First, certain basic *initial experiences of satisfaction* are phylogenetically adapted, that is, they are wired in by evolution. As Deutsch and Deutsch (1966) put it:

> It is the taste of water, the feeling of satiety, the sensations from the genitalia that an animal finds rewarding. The connection of these sensations with need reduction is not one which is made by each individual animal. Such a connection . . . has been made by the process of natural selection. Only those animals which have found certain sensations rewarding have survived. Learning . . . has already occurred in the species; the individual need not recapitulate it. (pp. 143-144)

Second, memory is required to record the experiences. Third, memories, when activated by any means, serve as a *goal*, which is to repeat the same experience of satisfaction (pleasure), i.e., to achieve *perceptual identity*. Freud (1900) even postulated that the initial activation might be an hallucinatory fulfillment, i.e., activation of the memory to hallucinatory intensity. Implausible as this may appear, Helen Keller (see Dahl 1965) vividly described just such hallucinatory memories of previous experiences of satisfaction (e.g., the taste of ice cream) during the period before she acquired language at about the age of six. The activation of memories is the attempt to achieve perceptual identity; until then the wish remains latent, that is, potential or descriptively and/or dynamically unconscious. Fourth, the inclusion of symbolic equivalence provides for the well known human capacity for finding and satisfying alternate wishes as substitutes for primary experiences of satisfaction. And last, lest we be limited to highly restricted and stereotyped behavior, it is necessary to assume, with good evidence to support the assumption (cf. Wolff 1966; Sroufe and Waters 1976; Nachman, Stern, and Best 1986), that novel experiences, for all their variety, are perceptually identical in the sense of being classified on the property of their unexpectedness and the aroused emotion of surprise. Novelty *qua* novelty is an intrinsic experience of satisfaction. This attraction to novelty assures a truly interesting creature, one with built-in opposing tendencies: on the one hand to repeat the same old experiences of satisfaction, but on the other to enlarge its repertoire, to satisfy its curiosity, and to expand its range of experiences. Whereas

---

[5]This definition is adapted from Freud's (1900) famous Seventh Chapter of the Interpretation of Dreams (cf. Dahl 1965, 1978, 1979, 1983), in which he proposed two very different models of motivation: one based on wishes, a very modern-looking cognitive model, and the other on instinctual energy. The instinctual energy model prevailed. How different psychological history might have been had the cognitive model survived instead!

once we were content with the taste of milk alone, we can eventually acquire an appetite for such odd tastes as a dry martini with onion.

Appetite

There are four essential structural components to an *appetite*:

(1) A *perception* of a specific internal (partly bodily) state, e.g., thirst or genital sensations,
(2) an *implicit wish* to reinstate (achieve perceptual identity with) a *previous experience of satisfaction*, e.g., the taste of water or the sensations from copulation,
(3) a *consummatory act*, e.g., drinking water or copulation,
(4) a *reafferent perception* of the feedback from the consummatory act, e.g., the taste of water or the genital sensations and their motor accompaniments, which eventually terminate the act.

Lorenz (1965) emphasized the learning that takes place in the context of appetitive behavior via the teaching function of phylogenetically adapted motor patterns interacting with "the reafference which the organism produces for itself by performing the consummatory act in the adequate consummatory situation" (p. 88). In other words, the teaching is accomplished by the feedback that terminates a consummatory act, i.e., that satisfies the wish. As Dahl (1978) wrote, "Using the model of infant feeding, we can say that the infant's consummatory act of sucking teaches it that the incoming fluid satisfies its appetites of hunger and thirst because that is the way the infant is built" (p. 389). And, we would now stress, built by natural selection in the course of evolution.

Appetites also have a number of conspicuous properties:

(1) *Peremptoriness*, i.e., they function as *instructions*.
(2) *Selectivity* of objects, i.e., there are objects that are specifically necessary for their satisfaction.
(3) *Displaceability* of objects, i.e., in the absence of the specific objects substitutes may suffice.
(4) A tendency to *self-stimulation* when satisfaction is possible.
(5) A tendency to *expansion of range* and *refinement of discrimination* of experiences that will satisfy—in other words a tendency to acquire new 'tastes.'

Two Major Functional Classes of Emotions

According to this theory, *Emotions* are a special class of appetites, exhibiting the same structure as somatic appetites plus all of the above properties of appetites. They function as *wishes* and *beliefs* in an evolutionarily given, phylogenetically adapted, nonverbal information feedback system, a system that is quintessentially *cognitive* in the sense of *knowing*. For the animal without symbolic language (which includes our evolutionary ancestors and every human before acquiring symbolic language) emotions are the primary *intelligence* system for surviving in a complex world of many dangers and for communicating and recognizing the intentions of members of ones own species.

The essential difference between emotions and somatic appetites lies in the fact that two major classes of emotions, the **IT** and the **ME** emotions, are specialized to fulfill different structural components of appetites. The **IT** emotions include the functions of the first three structural components: (1) the *perception of a specific intentional state*, (2) an implicit *wish* toward an object, and (3) a *consummatory act*. The **ME** emotions function as the fourth structural component of an appetite, namely (4) the *reafferent perception* of the feedback information about satisfaction or nonsatisfaction accompanying the consummatory act. Schwartz and Trabasso (1984) produced evidence for the psychological reality of the IT-ME distinction in a study which showed that 6-year-olds implicitly understand the classificatory dimensions as well as the implicit wishes and beliefs associated with the IT and ME emotions. We have the following definitions and examples:

IT Emotions have objects, function as *appetitive wishes* about those objects, and can be represented as: *P wishes that x*, where *x* is one of four formally definable classes of *consummatory acts*, defined by the intersection of two dimensions, valence (Attraction-Repulsion) and control (Active-Passive). Table 1 shows the four generic emotional appetites for objects, each with its generic wish and generic consummation and an arbitrary category number.

ME Emotions do not have objects, function as *beliefs*, and can be represented as: *P believes that y*, where *y* is information about the status of *satisfaction* or *nonsatisfaction* of appetitive and other significant wishes. The satisfaction of wishes results in the experience of *pleasure*, while nonsatisfaction results in the experience of *unpleasure*. Four generic classes are defined by the intersection of the two dimensions, valence (Positive-Negative) and

control (Active-Passive). Table 2 shows the four generic ME emotions, each with its generic belief, unique experience, goal, and typical name.

Generic Wish	Consummation	Sample Emotion
[1] Active Attraction (to) IT	Take Care of IT	Love
[2] Passive Attraction (from) IT	Explore IT	Surprise
[5] Active Repulsion (to) IT	Get Rid of IT	Anger
[6] Passive Repulsion (from) IT	Escape from IT	Fear

Table 1

Generic Belief	Experience (Goal)	Sample Emotion
[3] Passive Positive ME *Wishes have been satisfied*	PLEASURE	Contentment
[4] Active Positive ME *Wishes going well*	(Repeat)	Elation
[7] Passive Negative ME *Wishes can't be satisfied*	UNPLEASURE	Depression
[8] Active Negative ME *Wishes not going well*	(Get Rid of)	Anxiety

Table 2

Narrative	John's Emotions	John's Wishes & Beliefs
John sees Mary, an attractive classmate, at a party.	Interested, curious	**Wish:** [Cat 1-2] to get acquainted with Mary.
John feels attracted to Mary.	Affectionate, fascinated, friendly	**Wish:** [Cats 1&2] to consummate his several
John imagines what he and Mary will do; he walks	Eager, enthusiastic, bold, optimistic	**Belief**: [Cat 4] his wishes **will be** satisfied.
Mary suddenly turns to Fred, John's friend.	Anxious, distressed, upset, worried	**Belief:** [Cat 8] his wishes **may not be** satisfied.
Fred joins hands with Mary and they leave together.	Dejected, depressed, lonely,	**Belief:** [Cat 7] his wishes **can't be** satisfied.
John imagines that Fred has a serious accident.	Angry, jealous, envious	**Wish:** [Cat 5] to get rid of Fred.

Table 3

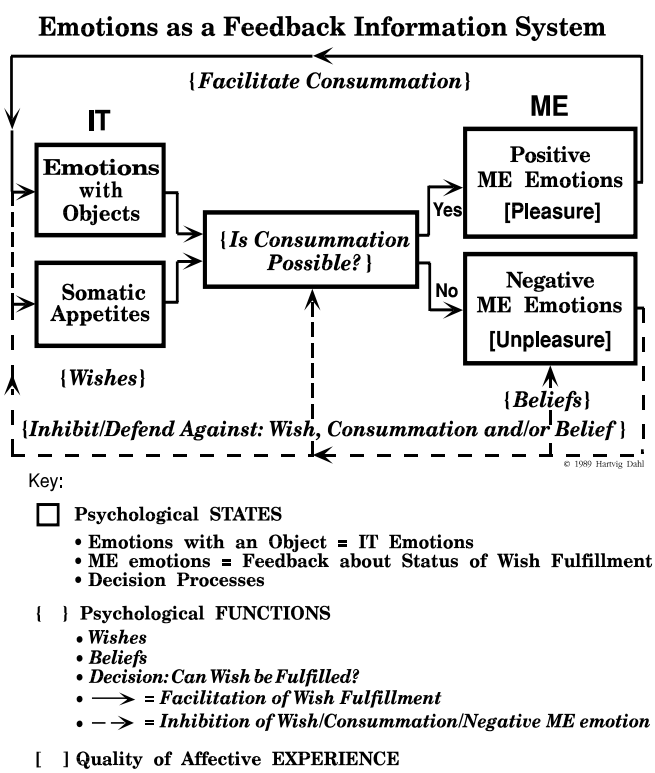

Figure 2

An adaptation of an unassuming little scenario created by Trabasso (1982), illustrates an everyday commonsense application of the theory. In Table 3, Column 1 tells a story about John, Column 2 lists some of John's emotions that one might plausibly infer from each event in the story and column 3 translates these emotion names into John's corresponding wishes and beliefs. The relationship between columns 1 and 2 is based on everyday commonsense knowledge and that between columns 2 and 3 is based on both the empirically established relationships (cf. Dahl and Stengel 1978) between emotion words and the classification dimensions and the definitions of the theory.

A Causal Feedback Model

In Figure 2 the question, "Is Consummation Possible?" involves a computation that is assumed to be made automatically following the activation of *any* significant wishes including the wishes implicit in the IT emotions. The outcome of this computation, *Yes* or *No*, determines the resulting major category of ME emotion. Many factors, historical and situational, influence the computation. Similarly, the consummatory act may or may not actually be carried out depending on many of the same factors. Part of a comprehensive program of research on emotions would have to include systematic investigation of these determinants.

The causal effect of the positive ME emotions is that of *facilitation* in the sense of helping to stabilize the memory of the experience of satisfaction, promoting both its reactivation under suitable conditions and fantasies about the objects

involved. Similarly, the causal effect of negative ME emotions is that of *inhibition* in the sense of a signal to invoke some kind of defense against: (1) the wish itself, that is, the activation of the memory of the experience of satisfaction, (2) the consummation of the wish, and/or (3) the negative ME emotion itself. Needless to say there is great variation in the success of such defenses, particularly against the negative emotion itself, often leading to auxiliary means such as alcohol and other drugs to get rid of the aversive quality of negative ME emotions.

Emotion Model and Commonsense Knowledge

One not necessarily self-evident advantage of these concepts lies in their representation of the cognitive content of emotions as wishes and beliefs.[6] Since emotions can be systematically represented as propositional attitudes, their propositional content can be incorporated into models and/or simulations of human cognitive processes, in particular, artificial intelligence models of cognition. Moreover one can systematically incorporate this model of the fundamental functions of emotions into intentional system theories that also use wishes and beliefs to predict and explain the behavior of complex bio-psycho-social systems. Both Dennett (1978, 1987, 1988) and Searle[7] (1983), for example, claim that emotions are typical intentional states, yet neither has proposed or borrowed a theory of emotions that accounts for the inclusion of emotional states as wishes and beliefs. We fully agree with Ortony and Turner (1990) that the question *Which are the basic emotions?* is misdirected. The emotion theory outlined here answers a different question: What are the *functions* of the different emotions that we can readily identify? And what wishes and beliefs do they entail?

Significance for AI

Historically AI researchers who have focused on building computational models of human cognitive processes have ignored a large part of human intelligence, namely, emotions and their functions. Consequently the emotional intelligence of humans has mistakenly been excluded from most models of human intelligent behavior. Among the few exceptions was Colby (1981), who incorporated a crude simulation of emotions in his computational model of a paranoid patient. "In the model, the strength of each affect is represented by a numerical value ranging from zero to 10" (p. 525). Another exception was Dyer (1983), who built commonsense knowledge of emotions into BORIS, a computer program designed to understand the emotions implicit in stories about everyday situations such as two people getting divorced. Their models, however, were not grounded in any theory of emotions.

Today AI is moving beyond simulation and modeling to computer systems that embody an understanding of human cognition and motivation. In this context the importance of emotions has finally been recognized. For example, according to Welty and Hoebel (1997), at the First International Conference on Autonomous Agents (Agents-97), "there were numerous calls for various new areas of Agent research. The keynote speaker called for a 'language of emotion . . .'," (p. 40). The role of emotions in creating lifelike animated characters was pointed out by Porter (1997):

> The underlying notion ... is that action is driven by cognitive processes in the character, that there is intelligence and personality and emotion. (p. 12) . . . [T]houghts and personality and emotions drive the characters. (p. 13)

A key question that an animator must consider is "*What is the central emotion with which to involve the audience?*"(p. 14). In the area of human-computer interaction, a computer system with a functional theory of emotion could construct realistic models of users' wishes and beliefs. In interactions with humans, such a system could understand and respond appropriately to users' intentions.

---

[6]Even, or perhaps especially, among emotion researchers there is still remarkable disagreement and unclarity over the cognition/emotion distinction. There are those who believe (e.g., Lazarus 1984, following Arnold 1960) that "appraisal" precedes emotional states and those who believe (e.g., Zajonc 1980, 1984) that "preferences need no inferences." Our position is that emotions are another given-by-evolution form of *knowing*, that is, of *cognizing*. Thus, for example, anger just *is* a computation by the nonverbal system whose generic knowledge (read *cognitive*) content is: a repulsion toward an object, and anxiety just *is* a computation whose generic cognitive content is: there is a probability that a relevant wish cannot be satisfied.

[7]Among 48 states that Searle (1983:4) listed as potential intentional states were 37 explicit emotion states (italicized here): belief, *fear, hope,* desire, *love, hate, aversion, liking, disliking, doubting,* wondering whether, *joy, elation, depression, anxiety, pride, remorse, sorrow, grief, guilt, rejoicing, irritation,* puzzlement, *acceptance, forgiveness, hostility, affection,* expectation, *anger, admiration, contempt, respect, indignation,* intention, wishing, wanting, imagining, fantasy, *shame, lust, disgust, animosity, terror, pleasure, abhorrence,* aspiration, *amusement,* and *disappointment*.

REFERENCES

Arnold, M. 1960. *Emotions and Personality, 1. Psychological Aspects.* New York: Columbia University Press.

Colby, K. 1981. Modeling a paranoid mind. *The Behavioral and Brain Sciences* 4:515-560.

Dahl, H. 1965. Observations on a "natural experiment": Helen Keller. *Journal of the American Psychoanalytic Association* 13:533-550.

Dahl, H. 1978. A new psychoanalytic model of motivation: emotions as appetites and messages. *Psychoanalysis and Contemporary Thought* 1:375-408.

Dahl, H. 1979. The appetite hypothesis of emotions: a new psychoanalytic model of motivation. In: C. E. Izard (Ed.) *Emotions in Personality and Psychopathology*, 201-225. New York: Plenum.

Dahl, H. 1983. On the definition and measurement of wishes. In: J. Masling (Ed.), *Empirical Studies of Psychoanalytical Theories*, 39-67. Hillsdale NJ: Analytic Press.

Dahl, H. 1995. An information feedback theory of emotions and defenses. In: H. Conte and R. Plutchik (Eds.) *Ego Defenses: Theory and Measurement*, 98-119. New York: Wiley.

Dahl, H. and Stengel, B. 1978. A classification of emotion words: a modification and partial test of de Rivera's decision theory of emotions. *Psychoanalysis and Contemporary Thought* 1:269-312.

Dahl, H., Hölzer, M. & Berry, J. 1992. *How to Classify Emotions for Psychotherapy Research.* Ulm: University of Ulm.

Darwin, C. 1872. *The Expression of the Emotions in Man and Animals.* London: John Murray.

Davitz, J. 1969) *The Language of Emotion*. New York: Academic Press.

Dennett, D. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology*, 3-22. Cambridge: MIT Press.

Dennett, D. 1987. *The Intentional Stance*. Cambridge: MIT Press.

Dennett, D. 1988. Précis of The Intentional Stance, *Brain and Behavioral Sciences* 11:495-546.

Dennett, D. 1996. Facing backwards on the problem of consciousness. *Journal of Consciousness Studies* 3:4-6.

de Rivera J. 1962. A decision theory of emotions. *Dissertation Abstracts International* (University Microfilm # 62-2356).

Deutsch, J. A. and Deutsch, D. 1966. *Physiological Psychology*. Homewood, IL: Dorsey Press.

Dyer, M. 1983. *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension*. Cambridge: MIT Press.

Ekman, P. (Ed.) 1973. *Darwin and Facial Expression: A Century of Research in Review*. New York: Academic Press.

Freud, S. 1900. The Interpretation of Dreams. *Standard Edition*, 4 & 5. London: Hogarth Press, 1953.

Freud, S. 1915. Instincts and their vicissitudes. *Standard Edition*, 14:117-140. London: Hogarth Press, 1957.

Heider, F. 1958. *The Psychology of Interpersonal Relations*. New York: John Wiley.

Izard, C. 1971. *The Face of Emotion*. New York: Appleton-Century-Crofts.

James. W. 1890. *The Principles of Psychology, Vol. 2*. New York: Dover, 1950.

Johnson-Laird, P. and Oatley, K. 1992. Basic emotions, rationality and folk theory. *Cognition and Emotion* 6:201-223.

Lazarus, R. 1984. On the primacy of cognition. *American Psychologist* 39: 124-129.

Lorenz, K. 1965. *Evolution and Modification of Behavior*. Chicago: University of Chicago Press.

McDougall, W. 1923. *Outline of Psychology.* New York: Scribners.

Minsky, M. 1975. A framework for representing knowledge. In: P. Winston (ed.) *The Psychology of Computer Vision*, 211-277. New York: McGraw-Hill.

Nachman, P, Stern, D., and Best, C. 1986. Affective reactions to stimuli and infants' preferences for novelty and familiarity. *Journal of the American Academy of Child Psychiatry* 25:801-804.

Ortony, A., Clore, G., and Foss, M. 1987. The referential structure of the affective lexicon. *Cognitive Science* 11:341-361.

Ortony, A. & Turner, T. 1990. What's basic about basic emotions? *Psychological Review* 97:315-331.

Porter, T. 1997. Creating lifelike characters in *Toy Story*. *SIGART Bulletin* 8:10-14.

Plutchik, R. 1962. *The Emotions: Facts, Theories and a New Model*. New York: Random House.

Pribram, K. and Melges, F. 1969. Psychophysiological basis of emotion. In: P. Vinken and G. Bruyn (Eds.) *Handbook of Clinical Neurology*, 316-342. Amsterdam: North-Holland.

Schwartz, R. and Trabasso, T. 1984. Children's understanding of emotions. In: Izard, C., Kagan, J. and Zajonc, R. (Eds.) *Emotions, Cognition and Behavior*, 409-437. Cambridge: Cambridge Univ. Press.

Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind.* New York: Cambridge.

Sroufe, L. and Waters, E. 1976. The ontogenesis of smiling and laughter: a perspective on the organization of development in infancy. *Psychological Review* 83:173-189.

Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief.* Cambridge: MIT Press.

Stone M. 1980. Modern concepts of emotion as prefigured in Descartes' "Passions of the Soul." *Journal of the American Academy of Psychoanalysis* 8:473-495.

Tomkins, S. 1970. Affect as the primary motivational system. In: M. Arnold (Ed.) *Feelings and Emotions: The Loyola Symposium*, 101-110. New York: Academic Press.

Trabasso, T. 1982. The importance of context in understanding discourse. In: R. Hogarth (Ed.) *Question Framing and Response Contingency: New Directions for Methodology of Social and Behavioral Sciences*, 77-89. San Francisco: Jossey-Bass.

Welty, C. and Hoebel, L. 1997. Backtracking. *SIGART Bulletin* 8:40.

Wolff, P. 1966. The causes, controls and organization of behavior in the neonate. *Psychological Issues*, 17. New York: International Universities Press.

Zajonc, R. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35:151-175.

Zajonc, R. 1984. On the primacy of affect. *American Psychologist* 39:117-123.