DE GRUYTER DOI 10.1515/jisys-2013-0069 — Journal of Intelligent Systems 2014; 23(3): 325–343

Sylvia Tidwell Scheuring and Arvin Agah*

An Emotion Theory Approach to Artificial Emotion Systems for Robots and Intelligent Systems: Survey and Classification

Abstract: To assist in the evaluation process when determining architectures for new robots and intelligent systems equipped with artificial emotions, it is beneficial to understand the systems that have been built previously. Other surveys have classified these systems on the basis of their technological features. In this survey paper, we present a classification system based on a model similar to that used in psychology and philosophy for theories of emotion. This makes possible a connection to thousands of years of discourse on the topic of emotion. Five theories of emotion are described based on an emotion theory model proposed by Power and Dalgleish. The paper provides classifications using a model of 10 new questions, for 14 major research projects that describe implementations or designs for systems that use artificial emotions for either robotics or general artificial intelligence. We also analyze the trends in the usage of various theories and complexity changes over time.

Keywords: Robot emotions, artificial emotion systems, emotions for artificial intelligence.

***Corresponding author: Arvin Agah,** Department of Electrical Engineering and Computer Science, University of Kansas, 1520 West 15th Street, Lawrence, KS 66045, USA, e-mail: agah@ku.edu

Sylvia Tidwell Scheuring: Department of Psychology and Research in Education, University of Kansas, Lawrence, KS, USA

1 Introduction

Rumbell et al. [20] proposed a classification of artificial emotion software systems using a computer scientist's approach to classification, based on four questions: "What is the action selection method (arbitration to command fusion)?"; "How is this architected (reactive to deliberative, symbolic to neural, continuous to discrete, hierarchical to distributed)?", "What roles do emotions serve (e.g., action selection, adaptation)?" and "Which emotional model is used (basic or dimensional)?" While the classification provides a foundation for making architectural decisions for robotic systems, restricting ourselves to such a technologically based classification scheme could limit our ability to predict how such systems correspond to human emotions and, therefore, the extent to which such systems will be convergent with the expectations of humans in an integrated artificial intelligence (AI) – or robot–human system. Such a technologically focused classification perspective may also limit our vision of how systems can take advantage of psychological research into the benefits of human emotions during decision making in complex environments.

We present a psychologically focused extension of the third question posed by Rumbell et al.: "What roles do emotions serve (e.g., action selection, adaptation)?" By classifying architectures using a model similar to how psychologists have classified theories of emotions throughout time, from Aristotle to the current era, constraints on the answers to Rumbell's other three questions can be identified.

An eight-question model was developed by Power and Dalgleish [18] to classify theories of human emotion. We propose a 10-question model for classification of robots and AI systems using artificial emotions, as shown in Table 1, side by side with the model offered by Power and Dalgleish.

By investigating robotics and AI systems with this Emotion Theory model classification scheme, we can relate the development of artificial emotions directly to a rich literature of more than 2000 years of theory on emotions in human systems. Examining how these systems conform to, or are distinguished from, various descriptions of human emotion may lead to paths for improvement of these systems, may help us understand

Table 1. Questions Asked by Power and Dalgleish [18] and the Emotion Theory Questions Used in this Paper for Classification.

Major Questions to Answer Regarding Emotion Theories [18]	Emotion Theory Classification Questions for Artificial Emotion Systems
1. What distinguishes an emotion from a non-emotion?	1. What distinguishes an emotion from a non-emotion within the system?
2. What are the constituent parts of an emotion, or are emotions irreducible?	2. What are the constituent parts of an emotion, or are emotions irreducible within the system?
3. What distinguishes one emotion from another?	3. Is there more than one emotion being used? If so, what distinguishes one emotion from another within the system?
4. What is the process of having an emotional experience?	4. What is the process of having an emotional experience for the system? Is it constant across emotions, or does it change for each emotion?
5. Why do we have emotions?	5. Why does the system have emotions? Do all emotions serve the same purposes?
6. What is the relationship between emotional states, moods, and temperament?	6. How do the system's emotional states differ from how humans characterize moods and temperament? Are emotions transient? If so, how long do they persist and what causes them to change?
7. How many emotions are there and what is the nature of their relationship with each other?	7. How many emotions does the system have and what is the nature of their relationship with each other? Do some emotions cause changes in other emotional states?
8. What is the difference between, and the relationship of, the so-called normal emotions and the emotional disorders?	8. Does the system detect and correct for the difference between, and the relationship of, the so-called normal emotions and the emotional disorders? And, if so, how does it do this?
	9. Where do the system's emotions originate? Are they created explicitly, learned from the environment, learned from social interactions, or some combination of these?
	10. Does the emotion improve the system's performance, and if so, in what way?

where such systems may or may not be compatible with future research in these areas in emotion theory and neural sciences, and may enable us to understand more about human emotions by providing a bridge back to using such systems as models for normal emotions or emotional disorders in humans.

Numerous technological methods exist, and new approaches are constantly being introduced. Because this classification scheme can be used for robotic and AI systems at any stage of development from business requirements specification, to system architecture, to deployment and maintenance, we believe it may provide a useful way of characterizing a system using artificial emotions in such a way that architects may more effectively weigh the benefits and disadvantages of existing and emerging technologies. The scheme may make it possible to distinguish technologies that work in general for artificial emotions from those that only work well when artificial emotions are characterized in specific ways.

2 Theories of Human Emotions

Given the importance of emotion to all aspects of human life, it should come as no surprise that the topic was discussed by many of the great philosophers – Plato, Aristotle, Spinoza, Hobbes, and Hume. More recently, emotions have become of interest to cognitive theorists as well. The philosophies of emotion have become connected with many other disciplines, including psychology, evolutionary biology, neuroscience, and robotics. It might be helpful to summarize a few of the theories of emotion as proposed by both the greats of philosophy and modern cognitive scientists to gain a better understanding of how these could be used in the design of artificial emotions. We will use the model of the eight questions proposed by Power and Dalgleish

[18] to facilitate comparison of the various theories. This section provides responses to each of the eight questions for five emotion theories, as well as the names of some of the theorists associated with those theories by those authors. We hope that we have not misrepresented the theories in our quest for brevity. A discussion of the possible technological implications of each of these theories is included. These may provide a starting place in the literature for technologists interested in extending their artificial emotion designs.

Feeling Theory: Power and Dalgleish's eight questions (Plato, Rene Descartes, William James, and, more recently, Damasio and Prinz):

1. For Plato, Descartes, and James, emotions are a part of the soul and not of the body, and are in opposition to rational thought. All feelings (including emotions) are a by-product of the movements of spirits within the pineal gland, or the more modern interpretation provided by Damasio and Prinz of emotions being the result of somatic changes in the body.
2. Emotions are a by-product of movements within the pineal gland. Emotions move us to the extent that the excitation stimulus can be harmful or profitable. Some emotions are mixtures of the six primary passions, which are irreducible.
3. They each have distinct motions of the bodily spirits.
4. An excitation factor (e.g., as seeing a large animal) causes changes in the soul, which results in spirit movements of which we become aware and perceive as emotion (e.g., fear). Emotions are not learned from the environment or experience; rather, they are intrinsic to the soul. Emotions directly cause actions (e.g., moving the legs to run from what is fearful). A later refinement of William James suggests "we feel … afraid because we tremble" [18]. Recent updates to this theory proposed by Damasio and Prinz suggest that somatic changes in the body (such as heart rate, blood pressure, and facial expression) cause changes in mental states that are what we think of as emotions [13].
5. Emotions (or the so-called passions) provide a defensive mechanism in the face of imperfections in the natural world.
6. These are not addressed.
7. There are six so-called primary passions: wonder, joy, sadness, love, hatred, and desire. Descartes "constructed a host of complex ones [emotions] out of these six" [21].
8. Plato saw emotional disorders as the result of lack of rational control or dominion over the emotions. According to Descartes, the will keeps the passions from overwhelming us and causing emotional disorders.

Subscribing to the Feeling Theory could have a number of technological implications. The Feeling Theory suggests that emotions directly cause actions. They are not cognitively processed first. This would mean that the actions of the AI or robotic system would depend directly on the emotional state. In the Feeling Theory, there are six emotions that are irreducible. From a technology perspective, these six emotions are binary. Complex emotions would be characterized by patterns of these binary variables. Emotional control in such systems would be done by suppression of the emotional expression, rather than substitution of expressions that differ from the internal emotional state of the system. Emotional disorders in such systems, if modeled, would be characterized by failure to suppress emotional reactions. Constraints to questions posed in Rumbell et al. include an action selection mechanism on the side of arbitration with emotional reaction defined by the pattern of emotions and rational suppression as an opposing force.

Behaviorist Theories of Emotion: Power and Dalgleish's eight questions (James Watson, B. F. Skinner, Gilbert Ryle):

1. Emotions are inherited patterns of reactions that involve systemic changes in the body, "particularly of the visceral and glandular systems" [18].
2. Emotions are irreducible.
3. Each emotion has a different systemic pattern.
4. A particular systemic bodily pattern is recognized by the brain. This pattern recognition is innate. This primes the individual experiencing the emotion to associate the particular environmental conditions with that emotion and seek to increase or decrease the probability of replicating the conditions now associated with that emotion.

5. Different emotions (which are inherited reactions to systemic bodily patterns) provide differentiated feedback to operant conditioning.
6. Temperament is an inherited sensitivity to detecting certain systemic bodily patterns. For example, some people are more or less inclined to detect the pattern that indicates fear. Moods are temporary inclinations to detect certain systemic bodily patterns. For example, after a particularly bad day, a person may be more sensitive to fear patterns than love patterns.
7. Fear, rage, and love (more akin to sexual drive).
8. The abnormal emotions would be the result of associations of many or most events with a particular systemic bodily pattern. For example, a person who is afraid of everything may be the result of poor environmental conditions interacting with an increased inherited sensitivity to the fear pattern.

Technologically speaking, each emotion from a Behaviorist point of view would be triggered by specific patterns of sensor readings or information. The presence of these emotional reactions remains constant. Full emotional control would require a system capable of avoidance of, attraction to, suppression of, and amplification of sensing patterns that result in particular emotional states. The fact that this emotional control of attention to sensor information is a function of prior emotional states means that the system must have a means of ignoring, amplifying, or moderating the awareness of sensory data based on their association with prior emotional states. This would suggest that the architecture would need to track emotion over time, associate sensory patterns with them, and adjust the processing of stimuli accordingly. The Behaviorist model of emotions would also suggest that emotions do not need to be learned from the environment; they can be encoded directly into the system.

Emotional disorders could be modeled by increasing the sensitivity of a system to certain emotions while suppressing others, which would result in overexpression of the favored emotional state even when sensor states would normally trigger other emotions. This suggests that systems that implement the Behaviorist model should learn the appropriate preferences for emotional states for a given environment.

Using Rumbell's model, architectures using the Behaviorist Theory would likely tend toward arbitration, reactive, symbolic, discrete, distributed models with basic emotions. Emotion roles supported would include motivation, attentional focus, alarm mechanisms, and action selection.

Evolutionary Approaches: Power and Dalgleish's eight questions (Robert Plutchick, Robert Frank, Charles Darwin, Paul Ekman):

1. Emotions are affect programs.
2. Emotions are irreducible; some are innate and universal, while others are learned from the environment.
3. Differences in affect distinguish one emotion from another.
4. Emotions are inherited reactions to specific situations. They are affect programs that prepare the individual to respond in a way most likely to support survival, dominance, mating, and effective affiliations. These behaviors are selected by natural selection. For example, if an organism responded to an attack with fear (running away or freezing in place) or rage (barring teeth), then the reaction most likely to support survival will be the programmed response retained by the species. Therefore, emotional experiences differ between individuals only where these goals will not be adversely affected by the behavior. Naturally occurring variations in the behaviors will be pruned by natural selection to the few that are most likely to promote the goals of survival, dominance, mating, and effective affiliations.
5. Emotional expressions serve evolutionary functions particularly in the roles of survival, dominance, mating, and affiliation.
6. This is not addressed.
7. Generally six are assumed to be universal: happiness, sadness, anger, surprise, and disgust, based on commonality of affect across cultures [9]. All other (infinitely many) emotions are assumed to be learned reactions to the environment and may therefore be different for each person. Plutchick claimed eight basic emotions and claimed that these are found in all organisms: fear/terror, anger/rage, joy/ecstasy, sadness/grief, acceptance/trust, disgust/loathing, expectancy/anticipation, and surprise/astonishment.
8. This is not addressed.

Fully adopting an Evolutionary approach to behaviors would require an architecture capable of generating new reactions to existing emotions. These reactions could be retained or suppressed on the basis of the results of the action. Ideally, additional systems with the same six basic emotions would be able to inherit these retained behaviors from other systems.

Each system would have the six or eight basic emotions; however, individual implementations would have an indefinite number of other emotions that are learned from the environment. Such indefinite behaviors could be modeled by non-semantic networks such as neural networks with stimuli as inputs and actions as outputs, with the specific structures and combination of weights in the network being termed "emotions." The complexity of these emotions could be modeled by the number of hidden layers or the number of nodes in each layer.

In Rumbell's model, an Evolutionary approach would suggest that the learned emotions would be neural rather than symbolic, continuous rather than discrete, reactive in real time, but deliberative in the decision to retain or discard emotion–reaction associations.

Functional and Cognitive Theories of Emotion: Power and Dalgleish's eight questions (Aristotle, Thomas Aquinas, Baruch Spinoza):

1. Emotions are defined by their roles within psychology. Emotions include reactive emotions such as anger, sadness, and fear, as well as the higher cognitive processes for such emotions as jealousy and envy.
2. Emotions are labels that we place on certain collections of action tendencies. In this way, specific action tendencies that make up a pattern of action tendencies could be thought of as the constituents of emotion.
3. The particular pattern of action tendencies for any given emotion for any given person is distinct from the action tendencies for another emotion. This makes the list of possible emotional states indefinite.
4. For Aristotle, emotions are contextual reactions to situations. An excitation object interacts with a current state of mind to produce a stimulus (an appraisal) that gives rise to an emotion, which then becomes an increased or decreased tendency to act in certain ways. For Thomas Aquinas, there is an initial impulse either to approach or avoid a particular stimulus that is a part of what Aristotle would call the state of mind. Some of these impulses (or primary emotions) are inherited (e.g., the impulse to avoid a shark). For Aquinas, there is a secondary cognitive process that can give rise to such emotions as fear or sorrow, which are secondary emotions. For Spinoza, the primary emotions are reflective and non-cognitive, and these are combined with a cognitive process of identification of causation to engender the more complex emotions. For example, love is the assignment of the cause of pleasure with the presence of a particular individual.
5. Emotions give us the capacity to act. They provide the action tendency for certain behaviors.
6. For Aristotle, moods and temperament would be a tendency to imbalance between emotional states.
7. Aristotle presented 10 specific emotions with two valences but did not propose that there were a limited number of emotions. The emotions he listed were as follows: positive valence emotions – calm, friendship, favor, pity – and negative valence emotions – anger, fear, shame, indignation, envy, and jealousy. Aristotle saw some emotions as the opposite of others (e.g., anger is the opposite of calm). Therefore, some emotions cannot co-occur with certain others. Thomas Aquinas would say that there are two primary emotions: attraction or avoidance, and that there are many more cognitively derived emotions. Spinoza would say that there are three primary emotions: desire, pleasure, or pain. For Spinoza then, emotions are anything that is a combination of these three primary emotions and an idea that associates these with a cause. Thus, "tickle" would be an emotion that associates a particular action of one's fingers with the result of pleasure.
8. For Aristotle, emotional disorders are the result of imbalance between reason and emotion or between emotions. Correction of an emotional disorder is done by rebalancing the experiences of emotion and reason.

Architecture employing a Functional or Cognitive theory of emotion would need to maintain three states: the current state of mind of the system, a state engendered by the pattern of sensory information, and the resulting emotional state. Some emotions would be directly connected to the pattern of sensory information, while

others (the cognitive emotions) would be the result of combining the state of mind with these patterns. The system may constrain certain emotions from co-occurrence. Emotional disorders could be modeled by inappropriate suppression of, or emphasis on, the cognitive emotions. This suggests that systems that implement a Functional or Cognitive model of emotions need to incorporate an additional layer that learns the conditions under which emotions should be primarily cognitive.

Constraints for Functional or Cognitive models using the system described by Rumbell et al. could include tending toward designs that include two or three subsystems: one that is reactive, symbolic, discrete, distributed, and has a basic emotional model; a second that is deliberative, neural, continuous, hierarchical, and has a dimensional emotional model; and possibly a third that learns to prioritize certain emotions or to give priority to one or the other of the previous two subsystems proactively, depending on expected environmental conditions.

Appraisal Theories – 20th Century Cognitive Theories of Emotion: Power and Dalgleish's eight questions (Richard Lazaryus, Nico Frijda, Klaus Scherer, Russell, Magda Arnold, Anthony Kenny, Erroll Bedford, Richard S. Peters, G. Pritcher, Robert Gordon, William Lyons):

1. Emotion is the label used to describe unusual physiological state change caused by particular appraisals.
2. Emotions are labels that we place on certain collections of physiological state changes that occur after particular appraisals. In this way, specific emotions could be thought of as consisting of patterns of physiological states in combination with appraisal patterns that immediately precede these physiological patterns. Some of these theorists have modeled emotions as positions in multidimensional space. Klaus Scherer proposed 18 dimensions. Russell proposed only two or three dimensions, e.g., arousal and valence.
3. Emotions are distinguished by their unique combinations of appraisals and physiological state changes.
4. For William Lyons, an external event results in attention, which causes an appraisal of the situation, which results in physiological change, which then causes a pattern of action tendencies that lead to particular behaviors happening with greater or lesser frequencies. Appraisals may be either conscious or unconscious. In some cases, both will occur, but one or the other may win out in any given situation, such as when a person consciously determines that spiders are no threat yet still unconsciously appraises a spider as dangerous and runs away.
5. Emotions are function in that they are used to select beneficial reactions to deal with specific situations.
6. Moods are short-term, either focused or unfocused, dispositions to particular appraisals. Temperament or personality would be considered a longer-term disposition to particular appraisals.
7. There are indefinitely many emotions. Some may be opposite to each other given that the accompanying physiological changes are a limiting factor (e.g., one cannot both raise and lower one's blood pressure simultaneously).
8. Emotional disorders are caused by conflicting cognitive appraisal that happen at different levels of the cognitive system. In some cases, one of the conflicting appraisals may be unconscious. One example of emotional disorders caused by this conflict is phobias.

Architectures employing an Appraisal Theory or the 20th Century Cognitive Model are more complex even than those for the Functional Theory. These architectures must maintain separate action selection processes simultaneously. Action can begin to occur on the basis of processes that complete first, but then be changed on the basis of processes that complete at a later time. These architectures can model abortive actions, such as moving toward something and then switching to another action plan. This means that when a new action plan is determined to be "better," the system must be capable of calculating the path from the current state toward the path intended by the new decision. Appraisal-based architectures could include three or four subsystems: the first three would be similar to those described under the Functional Theory, with the exception that the third system may predetermine action delays for one or the other of the first two subsystems; the fourth subsystem would facilitate reconciliation of partial actions executed with actions intended from later decisions.

3 Review of Published Works on Systems with Artificial Emotions

The published works were selected by using the phrases "robot emotion" and "artificial emotion" in Google Scholar. Initially, more than 50 papers were investigated on the basis of being in the first groups returned from each of these queries. Papers were eliminated if they did not describe implementations, simulations, or architectures of systems using artificial emotions. For each of the remaining works, the system described was classified by answering the 10 questions proposed earlier in this paper in Table 1. The results are included in Tables 2–15. References for additional reading include de Sousa [8], de Frietas and Queiroz [7], and Ziemke [26].

Arkin et al. [1] describe the implementation of Sony's AIBO and SDR entertainment robotics systems. The authors focus on two goals in the design of AIBO: to examine and model the behavior of creatures in their natural environment in order to model the natural, and therefore generally expected behaviors, and to incorporate a model of "motivational behavior" (e.g., emotions) that supports human interaction with the robotic system. The ethological model provides a range of behaviors from which to select, while the emotional model provides the mechanism that selects the appropriate behavior. The result is a robotic system that interacts with human companions in intuitive and engaging ways.

Kwon et al. [15] describe an intricate system that incorporates a complex model of emotions, implemented in an autonomous robotic platform. They describe an autonomous robot capable of recognizing emotional content in natural language and in touch gestures, e.g., the difference between a gentle and abrupt touch. The robot's own emotional state can be displayed in any of several modes, including facial expression – displayed on an on-board computer screen – gestures, and generated music. The emphasis in Kwon et al. is on emotional recognition and expression through multiple modalities, offering a wide range of experience both for the robot to recognize emotions in others, and for humans sensitive to different modes of expression to recognize the emotional state of the robot.

Breazeal and Brooks [3] describe the implementation of Kismet, a sociable although non-mobile robot capable of expressing a range of affect through control of facial actuators. Kismet contains a number of

Table 2. Classification of AIBO [1].

1	Emotions are modeled explicitly within the system, and are distinguished from ethological factors such as comfort seeking, play, and similar states.
2	The basic emotional states of the Ekman model are mapped into a three-dimensional space along axes of "pleasant," "arousal," and "confidence."
3	The Ekman model is used: there are six basic emotional states (happiness, anger, sadness, fear, surprise, and disgust). Each of these occupies a locus within the PAC (pleasant, arousal, confidence) space.
4	The current emotional state arises from the internal variables: nourishment, moisture, bladder distension, tiredness, curiosity, and affection. The internal variables are mapped to the PAC space, and the nearest emotional locus selected as the current emotional state. The selected emotion and the current drive (e.g., hunger) determine the specifics of behavior.
5	To select appropriate actions from the actions that fulfill a given drive, and express the robot's satisfaction with that fulfillment (or lack of it). All emotions serve this purpose.
6	Emotions change as rapidly as the internal variables determining emotional state. If the robot is denied the opportunity to fulfill a given drive, emotions can change very quickly. If the robot is currently "satisfied" with respect to its emotional needs, it may change only very slowly.
7	The system describes, explicitly, six emotional states, located at different positions in a three-space of pleasant, arousal, and confidence.
8	It is possible for the internal variables to fall outside "normal" ranges and produce behaviors that might be considered pathological, e.g., when affection is catastrophically low, the robot might ignore other drives in favor of seeking affection from its master. The system uses a set of covering variables to moderate such effects.
9	The system's emotions are created explicitly as labeled points in a three-space. The symbols that activate them, "MASTER," "FOOD," and so on, can be learned from the environment, but not the emotions themselves.
10	The system uses emotions to determine "correct" or appropriate behaviors in the context of interacting with human beings. Given that the design is for a companionate robot, the case could be made that the system would fail in its goal if this were not the case.

Table 3. Classification of Han Wool Robotics Company [15].

1	Emotions are distinct from sensors and apply separately to different persons. Others' emotions are perceived as distinct from the system's own emotions. Emotions are also distinct from the actions that indicate emotions.
2	There are no constituent parts, but they are relative, e.g., emotions are axes in a vector space, and the current emotional state is a location in that space.
3	There are multiple emotions. The largest-scale distinction is reactive/deliberative. The single reactive emotion is either present or absent; is categorical; and is one of happiness, anger, and fear. Different emotions are engendered by different sequences of percepts. The deliberative emotions are happiness, joy, sadness, distress, neutral, likeness, dislike, angry, hope, pride, fear, shame, and embarrassment.
4	Reacting to stimuli produces a quick, reactive emotion. Deliberative emotions exist in a viscous, springy, emotional space, the resultant vectors of which map to expressed emotions. The emotional experience is the same for all emotions: the distinction is in the path (e.g., reactive or deliberative) that leads to it.
5	To be able to express affect in communication with human beings. All emotions serve this purpose.
6	Emotions are characterized as existing in a thick, springy, fluid-filled space: they may be engendered by rule-based events (reactive emotions) or by deliberative modification. The viscosity could be compared to mood. The spring constant could be thought of as a temperament control. Emotions can change in response to various stimuli.
7	The article is unclear about the total number of emotions supported. The authors cite the OCC model, which provides 22 emotions, but lists only 13 of them. The emotions of this system are orthogonal and have little to do with one another, except for their mapping to an expression vector.
8	The issue of emotional disorders is not addressed.
9	Created explicitly from a set of rules. You do not learn to have emotions: you just have them.
10	Without them, the system would not work at all: it is all about the emotions in this case.

Table 4. Classification of KISMET [3].

1	Emotions are effectively discrete positions in "affect space." Elements of the system that do not arise from affect are not emotions. Note that this draws no distinction between an emotion and the expression of that emotion.
2	The emotions expressed by the system are locations in a three-dimensional affect space defined by the axes of stance (open, closed), valence (positive, negative), and arousal (low, high). Emotions per se are labeled loci within this space.
3	There are a potentially indefinite range of emotions used by the system, effectively limited only by the resolution of the actuators (e.g., servomotors) used to express them. Emotions are locations in an affect space, with integer values on each axis.
4	The emotional state of the system is tightly coupled to the cognitive state of the system, with various cognitive "releasers" activating drives and affective releasers, which in turn elicit emotions. Some parts of the system, e.g., the affective speech recognizer, activate the affect mechanism directly, rather than through the cognitive subsystem.
5	The system has emotions to "modulate the cognitive system of the robot to make it function better in a complex, unpredictable environment – to allow the robot to make better decisions, to learn more effectively, to interact more appropriately with others – than it could with its cognitive system alone." Emotional vectors directly control the position of actuators in the robot's face.
6	Emotions in the Kismet system are generally transient, in the sense that the system has no long-term ties to any particular emotion. However, this does not appear to be a limitation of the architecture per se, but of the experimental setup used by the authors as they were specifically interested in expression of affect.
7	The emotions explicitly labeled by the authors include anger, accepting, unhappy, sorrow, tired, calm, soothed, surprise, alert, joy, happy, disgust, fear, and stern. They are related through their locations in affect space, and indeed, can blend one into the other. Of these, anger, accepting, tired, unhappy, surprise, disgust, happy, fear, and stern are considered "basis postures" from which other affect states are interpolated. Shifting along a given axis does not trigger a change in position on another axis, e.g., changing stance does not trigger a change in valence or arousal. However, most shifts in emotion in the system occur on more than one axis simultaneously, presenting the appearance of related emotions.
8	The system described does not account for emotional disorders.
9	The emotions are explicitly preprogrammed reactions to the environment. If the same pattern of stimulus is provided, then the same emotion will occur. No environmental learning of emotions occurs.
10	The emotions allow for a large range of affect. Given that variety of affect was the goal of the system, it could be argued that it would not work without them.

cameras and a microphone; views from the cameras are combined to produce a stereoscopic sense of the robot's environment. Kismet's primary goal is to encourage humans to interact with it in various natural ways, especially focusing on the robot's ability to display its emotional state through facial expression and

Table 5. Classification of Samuel [4].

1	The robot head has three behavior elements: actions, emotions, and speech. Emotions are defined as collections of behaviors that are neither action nor speech.
2	Most of the emotions are irreducible, but to increase variety, sadness was divided into four types (guilt, sad, unhappy, and super grumpy)
3	Emotions in the system are determined by interactions. When a robot has not seen a face in a long time, sadness increases. When it has seen the same face for too long, the probability of boredom and irritation increases.
4	Emotions are generated by interaction periods with novel or less novel faces. Each emotion is expressed as a set of actuator positions or selection of particular audio files. The emotional behavior elements are in a list, each with a probability of occurrence, so that the behavior is not deterministic.
5	The robot has emotions to increase the interest in the interactions with it by people in a social setting.
6	Emotions appear to persist as long as the state that is causing them persists. They transition as soon as the conditions change. These appear to be emotions unconstrained by mood or temperament.
7	The paper describes 13 emotional states (neutral, sad, happy, scared, mean, blissful, irritated, lost, wicked, mad, uncertain, charming, and bored). It is unclear if irritated and bored are considered alternate states for the same emotion in the same way that sadness was modularized into four differing states for variety.
8	The paper does not discuss mechanisms for handling persistent states. It does mention that the robotic head can experience melancholy – "I'm only a robotic head, who cares of a robotic head?" – but it does not discuss if these states can get "stuck."
9	The system's emotional states are created explicitly on the basis of interaction parameters of new versus not new faces being recognized. States are probabilistic rather than deterministic but the probability values do not appear to be learned.
10	The basic task of the robot is to greet guests. The use of emotions appears to enhance the interest in the robot.

Table 6. Classification of Emotions in AGI [10].

1	Emotions are internal states that influence behavior. Emotions are interconnected with motivations. Motivation is defined as attraction to positive emotion and avoidance of negative emotion.
2	All specific emotions are combinations of the basic two emotions. Achievement of a goal results in the positive emotion "good." The presence of unexpected obstacles results in the negative emotion "bad." Therefore, complex emotions are learned from the environment as weighted combinations of positive and negative emotions.
3	There are two basic emotions and an unlimited number of composite emotions based on combinations of weights of these two basic emotions.
4	Emotions are constructed on the basis of classical conditioning. Emotional memories are constructed as associations between patterns of image or situations that resulted in achievement of a goal ("good") or detection of unexpected obstacles to meeting the goal "bad." The experience of "bad" emotions results in activation of planning.
5	This artificial general intelligence (AGI) architecture uses emotions to incorporate Isaac Asimov's three laws into an AGI system [2].
6	Emotions are based on aggregation of experiences into "good" and "bad" associations that persist in memory. So it is possible that they may persist if the emotional memory is weighted more than the current state analysis. The paper does not discuss this issue, but the design presented leaves a possible path to implementing emotions that are influenced by longer-term states such as mood or temperament.
7	The system has two basic emotions: positive and negative, and they are orthogonal. Both states can exist at the same time, and they do influence each other.
8	The paper acknowledges the possibility of disorders and considers these to be the possible consequent of poorly chosen training sets.
9	The system's emotions originate from the perception of "good" or "bad" situations.
10	In theory, the architecture increases the likelihood of friendly behavior so long as the trainings set used to train the robot is carefully chosen.

posture. For instance, if the robot is "sad," its ears will droop. If it is "angry," its brows will be drawn together, forming a collection of expressions easily interpretable by humans.

Chlebicki et al. [4] also present a system employing an implementation of a model head and conversational system. The authors describe two implementations of the mechanism of expression and a speech module that does not rely on truly synthesized speech, but on recorded voice files, from among which the system selects one or another for playback as part of the conversational process. Their initial implementation

Table 7. Classification of EMOBOT [11].

1	Emotions are internal states of the system. There are two types of internal states: drives and emotions. Drives are called "Primary Internal Values" because they are driven entirely from the sensory system. The drives are hunger, fatigue, homesickness, and curiosity. Emotions are dependent on the satisfaction of these drives. That is, emotions are a measure of how "out of balance" the drives are.
2	Emotions are dependent on drives and, in some cases, other emotions. In the implementation, they discussed each emotion could take on 256 different values.
3	Each emotion is defined by a particular pattern of weights of the distance of the drives from the preferred balanced condition; in this case, drives can have values from −1 to plus 1 with the balance condition being 0.
4	Emotions are experienced as a continuous function of the difference between the desired balance of each Primary Internal Value or drive. That is, emotions exist because drives are either satisfied or unsatisfied, or because other emotional states are in some particular configuration.
5	Emotion is intended to improve the system performance by providing an internal reinforcement signal that is used with other, external, reinforcement signals for unsupervised learning. All emotions serve this purpose. The system uses the state of the drives and makes action decisions learned using these reinforcements. There are four types of reinforcements: an objective function, human reinforcement signaling, trigger-based reinforcement, and reinforcement based on emotions.
6	Emotional states are dependent on previous emotional states and are described as decay or growth functions based on current states and previous emotional states. In this way, the emotional perspective of the system is somewhat persistent. This implies that the system would behave as if it has moods. It is possible that temperament could be supported in the system by allowing the weights to change for individual robots, but the authors do not describe this.
7	There are four emotions: fear, anger, boredom, and happiness, which roughly correspond to Damasio's [6] theories with boredom replacing sadness. Fear, anger, and boredom are distinct. Happiness is dependent on fear, anger, and sadness. All emotions are dependent on Primary Internal Values, so they somewhat interdependent.
8	The paper does not make reference to any emotional disorders.
9	Emotions are internal states of the system. There are two types of internal states: drives and emotions. Drives are called "Primary Internal Values" because they are driven entirely from the sensory system. The drives are hunger, fatigue, homesickness, and curiosity. Emotions are dependent on the satisfaction of these drives. That is, emotions are a measure of how "out of balance" the drives are.
10	The authors did not provide evidence that the emotions were beneficial in and of themselves, only that a system using them provides an interesting variety of behaviors.

dealt favorably with large social situations, interacting with crowds of visitors when presented at conferences in Wrocław and Vienna. The authors note the importance of the robot's expression of emotion when dealing with large numbers of humans.

Gavrilov [10] proposes a relatively simple model of emotional state determined by a neural network alongside what amounts to a prior knowledge base of fundamentally moral rules. He uses as his example the Three Laws of Robotics proposed by Isaac Asimov as a form of prior knowledge. The system he describes uses emotional state as a training input to a neural network, with the emotional state determined by conformance to the rules; if a situation is perceived as "breaking the rules," e.g., if a human is in danger, the system expresses this as negative emotion: it feels bad about the situation and uses this as a training input to the neural network that determines actions that will minimize bad feeling and maximize good feeling.

In the EMOBOT, Goerke [11] uses a hierarchical architecture consisting of a sensory upstream, an actuary downstream, and an internal value system/learning action controller. In the paper, he describes the progression of the internal values-based architecture through two simulations (the grid world EMOBOT and the real values EMOBOT) and then into a real-world implementation of a six-wheeled robot equipped with ultrasonic, infrared, and ambient light sensors. In each stage in the development, he has increased both the complexity of the behaviors and the internal value models. Goerke uses the terms "Drives" – for the states that depend directly on the pattern of sensory information – and "Emotions" – to describe states that are interrelated mathematical functions based on the degree of satisfaction (or lack thereof) of the drives. He uses these terms for their metaphorical value; he explicitly states that he makes no attempt to model the related psychological concepts.

Miwa et al. [16] provide a detailed account of the mechanism and emotion of an autonomous robot. Their robot incorporates visual, auditory, tactile, and olfactory senses, and is capable of a large range of

Table 8. Classification of WE-3RIV [16].

1	Emotions are distinguished from personalities. Personalities are defined as perceptual contexts under which emotional states will change and the direction of that change.
2	Emotions are regions in a three-dimensional space defined by the axes: pleasantness, arousal, and certainty.
3	There are a potentially indefinite number of emotions in the system. They are distinguished by their position in the three-dimensional emotion space.
4	Each situation causes changes in the position of the robot's emotion in the three-dimensional space. For example, for one of the personalities developed, losing sight of a target will cause a decrease in all three axes, whereas discovering a target will cause an increase in arousal, and certainty, but no change in pleasantness. All emotions are effected by sensory perception (visual, tactile, auditory, temperature, olfactory) or the lack of sensations at a given time. Personality will affect the expression of the mental state of the robot. Not all robots designed with this architecture will express all emotional states, even if they are experiencing them.
5	Emotions are used to increase the human-likeness of the interactions with people, and to make it possible to more closely match the personality of the robot to the humans with which it interacts.
6	Emotions change over time and with different perceptions. For example, if there is no stimulus, the arousal level decreases for at least one of the robots (WE-3RIV). If the same robot perceived alcohol, then arousal would decrease and pleasantness would increase while certainty remained the same, whereas the perception of smoke would increase arousal, decrease pleasantness, while still leaving certainty unchanged.
7	There are seven emotions: happiness, anger, disgust, fear, sadness, surprise, and neutral.
8	While the authors do not discuss mood or temperament, they do address the issue of how the architecture responds to differing personalities, which is similar to addressing temperament. The architecture separates the Equations of Emotions into those for the Sensing Personality and the Expression Personality. The Expression Personality is the degree to which the robot will express a given emotion that it is experiencing.
9	Emotions originate from a combination of Sensing Personality and perceptions. They are created using a table that defines each of the resulting changes in position of the emotion. Emotions are not learned from the environment, but Sensing and Expression Personality were modified by the authors to produce best-match emotional responses for interacting with particular people.
10	The authors stated that changes in the Sensing Personality and Expression Personality through their changes on the emotions experienced and expressed by the robot could increase the acceptability (or humanness) of their robot head.

Table 9. Classification of Modalin and Pudalin [19].

1	Emotions are defined as a "temporary state of mind, or affective state evoked by experiencing different situations."
2	Emotions are hierarchically organized into quadrants defined by the axes of valence and activation.
3	Emotions occupy different quadrants of the two-dimensional space (valence and activation), and are differentiated from each other by linguistic usage patterns.
4	Emotions of the user are detected using natural language processing. The context appropriateness of the emotion by the user is evaluated by comparison with a database constructed using emotive search terms in Google search. If the emotion is determined to be "normal" given the context, then the system provides emotional support to the user. If the emotion is considered unusual in the context, then support for emotional control is provided by the system.
5	The system does not itself experience emotions; rather, it acts as an emotional guide embedded in a conversation agent so that the agent can better guide the user toward the expression of contextually appropriate emotions using language selection and humor.
6	The system does not detect moods or temperament. Rather, it detects contextually appropriate emotional states and attempts to moderate emotional expression based on social normative expectations in each situational context. Thus, the system could be seen to counteract the effects of moods or extreme temperament or personality.
7	There are 10 emotions recognized and used (but not experienced) by the system: joy/delight, anger, sorrow/sadness, fear, shame/shyness/bashfulness, liking/fondness, dislike/detestation, excitement, relief, and surprise/amazement. They are categorized into four quadrants defined by the axes of valence and activation.
8	The system detects emotional disorders in the user and attempts to bring them back to emotional control. For example, if the user says, "I won the prize, but I feel so bored," the system will respond "You should be happy!" Or if the user says "I'd be happy if he died of cancer," the system will respond "Are you sure that is what you are feeling?"
9	The emotions were selected explicitly but their detection was learned using data mining techniques on natural language responses to Google searches. Emotion of the user is detected from a system that used social interaction as a training technique.
10	The system provides a promising path improved by understanding the appropriateness of emotional reactions to given situations and then changing the response behaviors to support social norms. Detection of interactions based on detection and classification of user's emotional states as normative/non-normative given context appears to increase expression by users of positive emotional states in conversations.

Table 10. Classification of SMAINE, Poppy, Prudence, Spike, and Obadiah [22].

1	Emotion is that which helps to select socially appropriate multimodal behaviors.
2	There is an implication that emotions within the system are based on axes (valence and activity): cheerful is positive–active, pragmatic is positive–inactive, aggressive is negative–active, and gloomy is negative–inactive.
3	The valence and activity appear to be what distinguish one emotion from the others.
4	The system detects the emotional state of the user from facial expression, body language processing, voice and natural language processing, and chooses utterances, facial expressions, and body language that will guide the user toward its own emotional state. For example, Poppy (the cheerful robot) will nod and smile and say "Yeah" to a positive user emotion, whereas Obadiah (the gloomy robot) will either not respond or will frown or shake his head in disagreement.
5	The emotions in the system are designed to draw the user toward the agent's emotional state. The goal of the system is to act naturally (like a human), make plausible action selections for both verbal and non-verbal actions, have a consistent personality, and interact as if it is aware of the user.
6	Each of the robots has a fixed personality (i.e., emotional state).
7	The system can support any of four emotional states but does not appear to support changes in emotional state.
8	The system does not address emotional disorders.
9	The emotions that the system has appear to be hard coded (explicit). Recognition of emotions expressed by the user is learned from the user behavior.
10	The authors propose a method for evaluating the success of emotion-based interaction systems, but they do not provide results for their system.

Table 11. Mathematical Classification of Action Tendency [23].

1	Emotions are only associated with objects, and are triggered by events and expectations. Emotions modify the probability of actions so as to determine the set of actions to perform in a given situation. A mathematical model of emotions is provided by the authors.
2	Emotions are classified in a hierarchical structure by valence. That is, emotions can be positive or negative. Emotions decay differentially based on this valence (negative emotions decay faster than positive ones).
3	Emotions are differentiated by the functional goals, action tendencies, and end-state goals. Functional goals: consume, readiness, control, protection, caution, orientation, recuperation. Action tendencies: approach, free action, agnostic, avoidance, attending, rejecting, inhibition, inactivity. End-state goals: access, obstruction removed, own inaccessibility, identification, object removed, absence of response.
4	Emotional experiences are the result of events or variance from expectations. Both positive and negative emotions can be triggered. Emotions decay over time after having been triggered, so emotions that are currently being experienced may have been triggered at different times. The goal of the system is to minimize the intensity of any negative emotions that it may be experiencing. Actions are first pruned on the basis that they are expected to result in progress toward the goal, and then the remaining set of actions are pruned on the basis of supporting an emotional goal. Idling may be selected as an action most likely to support all goals (e.g., counting to 10 before acting when angry). Positive emotions tend to lead to proceeding as originally planned, whereas negative emotions tend to cause changes in plans (reconsideration).
5	The system has emotions to prune the search space of possible actions based on past and current emotional context.
6	Emotions are triggered by events, but emotional states are the sum of all emotional states that have been triggered, and as emotions can persist (but decay) over time, moods could be thought of as the persistent emotional state that has not been changed by sufficient triggering of alternate emotional states. The system could account for temperament or personality by modifying what the authors call "overall" action tendency.
7	There are 22 emotional states based on the Ortony, Clore, and Collins (OCC) model [17]: gratification, gratitude, pride, admiration, joy, happy for, gloating, hope, satisfaction, relief, love, remorse, anger, shame, reproach, distress, resentment, pity, fear, disappointment, fears-confirmed, hate.
8	The authors do not address this directly, but their architecture could be used to explore these issues, as the expression and decay of emotional triggers are mathematical in nature and they have a concept of overall action tendency that could be used to model emotional disorders.
9	The authors do not address this directly, but their architecture could be used to explore these issues, as the expression and decay of emotional triggers are mathematical in nature and they have a concept of overall action tendency that could be used to model emotional disorders.
10	The authors assume that emotions can provide for the possibility of better and faster action selection by pruning the action selection space, but they do not provide any experimental evidence to support this assumption.

Table 12. Classification of Emotion and KANSEI [24].

1	Emotions are that which is used to select actions for the robot. Emotions are the only things used as the basis for action selection in this system. In this way, emotions are distinct from actions.
2	Emotions are irreducible in the system.
3	Emotions are distinguished by the differences in the actions that they engender.
4	Emotions of a user are detected by recognition of user gestures (learned using a neural network). The robot then takes on the same emotional state as the user and uses this emotion to select a particular combination of movements, lights, and music that reflect that emotional state.
5	Emotions are the primary form of communication. Emotions are understood and expressed.
6	All emotions are always present, but only the perceived dominant emotion of the user is the one expressed. The dominant emotion changes as the system detects changes in the emotional state of the user. The system does not appear to have any concept of persistence of emotional state, so neither mood nor temperament is addressed.
7	There are four emotions: happy, angry, tranquil, and melancholy. Emotions compete with each other and exist in an emotional space. Therefore, if one emotion is occupying more space (wide), the others are forced to occupy less (narrow) through competition in a neural network.
8	This system does not consider emotional disorders.
9	The dominant emotions are explicitly defined reactions to specific stimuli. They change based on the detection of these stimuli. All emotions are always present in the system, but one becomes dominant at any given time based on the probabilistic perception of the associated stimuli.
10	The system's only function is the perception and multimodal expression of emotion, so without them the system does not function.

Table 13. Classification of LIDA [25].

1	"Emotions are feelings with cognitive content." Non-emotional feelings, in contrast, are feelings without cognitive content (e.g., thirst or the pain from a scratch).
2	Feelings (both emotional and non-emotional) are hierarchically organized into categories: those with positive or negative valence.
3	Multiple emotions can be present simultaneously in the system. Each emotion is separate from the others, but their activation level can change over time.
4	Emotions are a subset of feelings in the LIDA system. Feelings that have cognitive content are considered emotions. Emotions are a consequence of sensory perception colored by previous emotional experiences. Emotions change the structure of memory in the system. These memories then activate attention to future events and amplify or suppress through spreading activation processes the creation and retention of like emotions and associated memories.
5	The overall purpose of the system is to provide a basis for moral decision making in an artificial general intelligence (AGI). Emotions are used as cues into transient and episodic memories. These activations help the memories to activate and become part of consciousness. Emotions that occur in the context of a scheme (as defined in psychology) also increase activation of procedural memory. Thus, emotions engender motivation to perform a particular action. Higher affect levels (up to a point) result in greater learning until affect becomes too high, at which point the affect begins to impair learning. Feelings also bias the specifics of performance of the action. For example, "an angry person picking up a soda may squeeze it harder than if he weren't angry."
6	Moods could be modeled in the system as co-activation of situations with certain emotional memories in either the episodic or procedural memory. Selective activation of positive or negative valenced emotions could be done using LIDA as a model for temperament, but the authors do not discuss this.
7	As feeling nodes can be created at will, there is no limit to the number of emotions that the system can represent. Each node can cause activation of related nodes, so emotions can affect the activation level of other emotions, essentially increasing the probability that these emotions will become conscious and therefore affect learning and action selection.
8	Disorders could be represented in LIDA by inhibiting the decay of certain emotions in the system, but the authors do not discuss this.
9	The authors do not discuss how the new nodes in the activation network are defined, so it is unclear what would cause emotions in the system.
10	The authors argue that emotions improve the likelihood of selection of "moral" actions in the time allotted for action selection, but they do not provide experimental evidence to support this theory.

Table 14. Classification of EBDI [12].

1	Emotions are distinguished from beliefs, desires, and intentions. A belief is a proposition regarding the true state of the world. An intention is a goal – something that you want to be true about the world. Desires are the methods that are to be used to fulfill intentions. Emotions are perspectives on other agents (e.g., thankful/hate).
2	Emotions in this system are irreducible.
3	There are two emotions: hate and thankful. They are disguised by the conditions that cause them. Hate is caused when an agent is given false information by another agent. Thankfulness is engendered when an agent is given correct information.
4	Both emotions are caused by the evaluation of truth or oppositions shared with the agent by other agents. For example, in Tileworld, the agent is told that a tile is in a given location by the other agent. If this is true, then the resulting emotion is thankfulness. If this is not true, then the resulting emotion is hate. All emotions are directed at other agents and control later communications with those agents.
5	Emotions are used to control the belief in future information from a given source and to determine whether or not to lie to the other agent.
6	Emotions are changed by an evaluation of the truth of a proposition shared by another agent. They are dependent on the previous emotion regarding that agent. That is, they are incrementally changed by interactions rather than being entirely controlled by the last interaction. Thankfulness increases with each correct proposition and decreases with each incorrect proposition. Hate increases with each incorrect, and decreases with each correct proposition.
7	There are two emotions, and they are opposites. If hate increases, then thankfulness decreases and vice versa.
8	This system does not consider emotional disorders.
9	The emotions are explicitly coded responses to specific situations.
10	The authors claim that emotion increases the performance of the system over agents without emotions. They referenced data comparing the performance of agents with and without emotions in Tileworld and showed that those who could be thankful or hate other agents were more successful at covering tile holes in a world in which there were two other agents: a liar and a truth teller, each without such emotions.

Table 15. Classification of Silbot [14].

1	Emotions are multimodal expressions (facial expressions + body language)
2	Emotions are hierarchical; grouped into positive emotions, neutral emotions, and negative emotions.
3	The architecture supports a nearly infinite number of emotional expressions, but only 128 are supported in their implementation. Names are not provided for these 128 in their paper.
4	The process of changing the system's emotion is not defined, but the process of defining the expression of the emotion is. Each emotion is associated in the database with a particular sequence of facial expressions, arm movements, neck movements, and wheel movements.
5	Emotions are used to facilitate multimodal expressions. That is, the system uses the emotional state of the robot in combination with the specific utterances to be made to determine which permutations of facial expression, neck movement, and arm and wheel movements should accompany the utterance and their timing with regard to the utterance.
6	There is insufficient information in the paper to determine how these emotions are related to mood and temperament.
7	There is no limitation in the paper on the number of emotions; in fact, the emotions are not mentioned in particular. The authors are instead describing an architecture for implementing multimodal expression of any set of emotions.
8	This system does not consider emotional disorders.
9	The authors do not discuss where the emotions originate. They only discuss their usage in determining expressions.
10	Emotions improve the system's performance by enabling rich contextual facial expressions and body language to the spoken utterances.

expression, incorporating mechanical actuators for the position of various components of its face, and a thin electroluminescent sheet allowing it to change "skin" color slightly, when expressing, for example, anger or embarrassment. The robot's personality is based on a separable Sensing Personality and Expression Personality, with both personalities coordinating through a mental model that determines overall emotional state as a position within an emotional space determined by axes of pleasantness, uncertainty, and activation.

Ptaszynski et al. [19] focus on the goal of the interaction between a human and a conversational agent. The agent attempts to understand the emotional expressions of its human conversational partner and determine whether those emotions are appropriate to the context; it can then choose whether to be directly supportive of its human partner, or suggest to its partner that the emotion it has expressed is inappropriate to the circumstances. Determination of whether an emotional expression is appropriate, convergent with social

norms, is accomplished through a data store of emotional affect terms drawn from "mining" publically available documents found on the Web. Using different reactive models, e.g., task-focused or humor-focused, Ptaszynski and his colleagues were able to show stronger positive reactions in the human conversational partner to some models than to others, e.g., humans reacted more favorable to models that made their point with humor.

Schröder and McKeown [22] describe a European project that is focused on the development of a system with "social and emotional intelligence," or what they term a Sensitive Artificial Listener. The paper describes the implementation and behaviors of four different personalities that were tested (Poppy, Spike, Obadiah, and Prudence), each with a different affect focus (cheerful, aggressive, gloomy, or pragmatic). The artificial personalities react to listeners with both body language (such as nodding), facial expression, and a narrow set of utterances that are intended to appear consistent with their personality, and designed to steer the user's affect toward that of the agent. This system's goals are convergent with those of the conversational agent described by Ptaszynski et al., but place less emphasis on verbal or written language and more on the non-linguistic cues used to affect emotional change.

Steunebrink et al. [23] present a mathematical operation-based model of emotion, emphasizing the role that emotions play in behavior at the level of action tendency; that is, the effect that emotions have on the set of actions from which the agent chooses. Emotions here are presented in a formal mathematical model derived from the work of Ortony et al. [17], describing emotions in terms of objects, actions, and events. While the authors do not describe an implementation of their method, unlike many researchers in the field of artificial emotions, their model incorporates a very wide range of actions in response to emotional conditions, such as idling, the "take a deep breath and count to ten" reaction to negative emotion. The range of actions they model also takes into account reconsidering the agent's actions in light of new information or changing state, and the option of uncommitting to an action, and replanning, based on feeling differently about the goal.

Suzuki et al. [24] describe a small unobtrusive wheeled robot equipped with camera and ultrasonic sensors that can sense the dominant emotions of the human companions, and is designed to increase their creativity through reflecting the perceived dominant emotional state of its human companions. The authors make explicit reference to using modern psychological theories to inspire their design of self-adaption of competing emotional states resulting in one single dominant expressed emotion at any given time. Detected human gestures and movement are mapped onto emotional states of the robot, which are then expressed as styles of movement, robot behavior, visual media, environmental lighting control, and music. Similar to both Schröder and McKeown and Ptaszynski et al., Suzuki et al. rely on an emotional architecture to influence the internal emotional state of humans. In this case, the authors employ the mobility of robotics and an emotional architecture to inspire human creativity through the establishment of emotionally supportive environments.

Wallach et al. [25] postulate that feelings, rules, and virtues can be used to extend the artificial general intelligence (AGI) model, LIDA, to incorporate higher-order cognitive tasks such as those of Machine Morality. In doing so, the authors provide background on the general state of AGI and the details of the LIDA architecture. While they do not describe a system that has been implemented, they suggest that LIDA, because of its ability to support multiple inputs, may provide an ideal platform for modules that convert sensory information to emotional responses constrained by moral values. Their proposed architecture, with a rich multidimensional emotions that are retained over time, affect future perception and emotional interpretation, and are learned in the context of societal values, maps more closely to modern psychological theories than the other architectures reviewed herein.

Jiang et al. [12] have implemented a model of emotional agents in a simulated "Tileworld," where virtual agents attempt to cover holes with tiles. Multiple agents in the world can communicate information about the locations of tiles and holes. In the model presented by the authors, some agents can learn to trust and cooperate with other agents, or "lie" in their communications, and so become untrustworthy. They are able to show that agents that track how they "feel" about specific other agents are more successful in their tasks than agents that do not incorporate an emotional model.

Kim et al. [14] describe an autonomous robot, called Silbot, intended as a "mascot," or companion for the elderly. Silbot is capable of expressing emotion in multiple modes, including word choice, facial expression,

and posture, the latter category including arm position, neck position, and wheel movement. The basic library of gestures and expressions are developed by motion capture of professional actors, rather than generated from somatic principles. Terms that the robot intends to communicate are filtered through an emotional layer, which provides specific word-choice selections and settings for motor components that are intended to communicate the robot's current emotional state.

4 Grouping the Architectures

We considered various groupings of the presented papers. Ordering by date would make it possible to demonstrate the tendency toward complexity of implementations that we saw when investigating the various architectures. Grouping by the type of implementation would make it easier for architects building different types of artificial emotion systems to see which papers were most central to their area of interest. Ordering by the most closely associated emotional theory would show which theories have been implemented by the various teams and which have been neglected by authors to date. We determined that it would be best to provide all these groupings to allow researchers to select the grouping or ordering that best fits their current needs.

4.1 Time Order of Publication and Complexity

To evaluate the complexity of the works, the complexity rubric is defined in Table 16. As shown in Table 17, both the complexity of the designs and the frequency of publication of systems have increased over time. Half of the papers that were identified were published from 2008 to 2010.

4.2 Types of Implementations

Artificial emotions have been implemented for both robotics and AI systems. There was about an even split between papers addressing AI and those addressing robotics with the AI systems tending toward more complex implementations (Table 17).

4.3 Most Closely Related Emotional Theory

In general, it is not important to distinguish between the theories discussed by the authors and the most closely associated theories. Many papers that use Ekman's research on emotional expression have imple-

Table 16. Complexity Rubric.

Complexity	Number of Emotions	Emotions	Emotional Associations	Reactions
Very low	One dimension or two or fewer distinct emotions		All preprogrammed	
Low	Two dimensions or three to six or distinct emotions		All preprogrammed	
Low–moderate	Three dimensions or three to six but less than indefinite number of distinct emotions		All preprogrammed	
Moderate	Three dimensions or three to six but less than indefinite number of distinct emotions		One learned/two preprogrammed	
Moderate–high	Three dimensions or three to six but less than indefinite number of distinct emotions		Two learned/one preprogrammed	
High	Four or more dimensions or more than six but less than an indefinite number of distinct emotions or all learned			
Very high	Four or more dimensions or an indefinite number of distinct emotions and all learned			

Table 17. Approaches Classified by Complexity, Domain, and Emotion Theory.

Year	Work	Complexity of Implementation or Design	Domain	Emotion Theory
1998	Suzuki et al. [24]	Low–moderate	Robotics	Behaviorist
2001	Miwa et al. [16]	Low–moderate	Robotics	Functional and Cognitive
2003	Arkin et al. [1]	Moderate	Robotics	Feeling Theory
2005	Breazeal and Brooks [3]	Low–moderate	Robotics	Functional and Cognitive
2006	Goerke [11]	Low	Robotics	Feeling Theory
2007	Jiang et al. [12]	Very low	Artificial intelligence	Functional and Cognitive
2007	Kwon et al. [15]	High	Robotics	Functional and Cognitive
2008	Gavrilov [10]	Very low	Artificial intelligence	Behaviorist
2009	Steunebrink et al. [23]	High	Artificial intelligence	Appraisal Theories
2009	Ptaszynski et al. [19]	High	Artificial intelligence	Functional and Cognitive
2010	Chlebicki et al. [4]	High	Robotics	Evolutionary
2010	Kim et al. [14]	High	Robotics	Evolutionary
2010	Schröder and McKeown [22]	Low	Artificial intelligence	Functional and Cognitive
2010	Wallach et al. [25]	Very high	Artificial intelligence	Appraisal Theories

mentations most closely associated with theories other than Evolutionary theories. All five of the theories we discussed at the beginning of this paper have been implemented in at least two of the reviewed articles. The most popular theory of emotion is the Functional and Cognitive theories originating with Aristotle, which were implemented in nearly half of the papers reviewed. The remaining four theories were each associated with two of the papers (Table 17).

5 Conclusion

Christian [5] describes his preparation for being a confederate at the Loebner Prize competition for the Most Human Computer. The confederates are the human beings that act as the decoys for the computers in this international competition held each year. The computer system that is able to gain the most "human" votes when compared with these confederates is awarded the "Most Human Computer" award for that year. The descriptions Christian provides of the historical test constraints as well as the current constraints of the test, and of the systems that have won make it clear that, at least in the case of the Loebner Prize, we no longer consider rationality to be the only thing that defines "intelligence" in our AI systems.

In fact, the humans gaining the most "computer" votes are generally those with the highest degree of rationality and crystallized knowledge. The computer systems that have won the prize in most recent years, however, either used recorded bits of human text conversations or used systems that behaved as if they were abusive humans or emotionally obsessive ones. This would indicate that the "humanness" sought by the Turing test has shifted in recent years from meaning "rational" to meaning "emotionally rational," with an emphasis on the emotional.

Despite some authors' interpretations of the Turing test as a test of rational behavior (e.g., Schroder and McKeown [22]), the most famous implementation of that test on computer systems is no longer a test of rational thought alone. While historically, there were constraints in place that gave the advantage to "rational" systems, current human judges for the Loebner Prize, focusing on the capabilities and limitations of current AI systems, are pushing the boundary of the definition "intelligence" more toward an expectation of fully human behaviors.

Users and designers are increasingly coming to see AI systems as social systems (e.g., Schröder and McKeown [22]). Understanding how previous emotion-based AI systems have been designed, and connecting these with the psychology literature, may provide a bridge that future developers can use to pass these evolving definitions of the Turing test.

The systems we have examined in this paper use various techniques, including supervised and unsupervised machine learning, data mining, fuzzy logic, Bayesian inference networks, neural networks, statistical analysis, natural language processing, and behavioral tagging. These are used to recognize, construct, or use human-like emotions to improve system performance in both interacting with humans and other agents and choosing effective plans or actions. Each of these conceptualizes emotions in their own way. We have attempted to focus our classification of these systems on the way past designers have viewed emotions and the purpose for which emotions are used. This may help future developers of robots and AI systems to decide how to conceptualize emotions and their roles for new systems.

Received September 7, 2013; previously published online February 7, 2014.

Bibliography

[1] R. C. Arkin, M. Fujita, T. Takagi and R. Hasegawa, An ethological and emotional basis for human-robot interaction, *Robot. Auton. Syst.* **42** (2003), 191.
[2] I. Asimov, *I, robot*, Doubleday & Company, New York, 1950.
[3] C. Breazeal and R. Brooks, Robot emotion: a functional perspective, in: *Who Needs Emotions?*, J.-M. Fellous and M. A. Arbib. (eds.), Oxford University Press, New York, 2005.
[4] S. Chlebicki, J. Kedzierski and M. Żarkowski, Control system for a social robot, Scientific Publications of Warsaw University of Technology, **II** (2010), 713–722.
[5] B. Christian, *The most human human: what talking with computers teaches us about what it means to be alive*, Doubleday, New York, 2011.
[6] A. R. Damasio, *Descartes' error: emotion, reason and the human brain robot*, Picador, New York, 1994.
[7] J. S. de Frietas and J. Queiroz, Artificial emotions: are we ready for them?, *Lect. Notes Comput. Sci.* **4648** (2007), 223–232.
[8] R. de Sousa, Emotion, in: *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.) (2012). http://plato.stanford.edu/archives/spr2012/entries/emotion/.
[9] P. Ekman, E. R. Sorenson and W. V. Friesen, Pan-cultural elements in facial displays of emotion, *Science* **164** (1969), 86–88.
[10] A. Gavrilov, Emotions and a prior knowledge representation in general artificial intelligence, in: *Intelligent Information and Engineering Systems INFOS 2008*, Varna, Bulgaria, 2008.
[11] N. Goerke, EMOBOT: a robot control architecture based on emotion-like internal values, in: *Mobile Robots, Moving Intelligence*, J. Buchli (ed.), pp. 75–94, ARS/pIV, Germany, 2006.
[12] H. Jiang, J. M. Vidal and M. N. Huhns, EBDI: an architecture for emotional agents, in: *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS '07)*, pp. 38–40, ACM Press, Honolulu, HI, 2007.
[13] G. Johnson, Theories of emotion, in: *Internet Encyclopaedia of Philosophy: A Peer-Reviewed Academic Resource*, J. Fieser and B. Dowden (eds.) (2009). www.iep.utm.edu/emotion.
[14] W. H. Kim, J. W. Park, W. H. Lee and M. J. Chung, Hierarchical database based on feature parameters for various multimodal expression generation of robot, in: *2010 IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO)*, Seoul, Korea, 2010.
[15] D.-S. Kwon, Y. K. Kwak, J. C. Park, M. J. Chung, E.-S. Jee, K.-S. Park, H.-R. Kim, Y.-M. Kim, J.-C. Park, E. H. Kim, K. H. Hyun, H.-J. Min, H. S. Lee, J. W. Park, S. H. Jo, S.-Y. Park and K.-W. Lee, Emotion interaction system for a service robot, in: *16th IEEE International Conference on Robot and Human Interactive Communication*, Jeju, Korea, 2007.
[16] H. Miwa, T. Umetsu, A. Takanishi and H. Takanobu, Robot personality based on the equations of emotion defined in the 3D mental space, in: *Proceedings of the 2001 IEEE International Conference on Robotics and Automation*, Seoul, Korea, 2001.
[17] A. Ortony, G. L. Clore and A. Collins, *The cognitive structure of emotions*, Cambridge University Press, Cambridge, England, 1988.
[18] M. Power and T. Dalgleish, *Cognition and emotion: from order to disorder*, 2nd ed., Psychology Press Taylor & Francis Group, Hove, 2008.
[19] M. Ptaszynski, P. Bybala, W. Shi, R. Rzepka and K. Araki, Towards context aware emotional intelligence in machines: computing contextual appropriateness of affective states, in: *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, Pasadena, CA, 2009.
[20] T. Rumbell, J. Barnden, S. Denham and T. Wennekers, Emotions in autonomous agents: comparative analysis of mechanisms and functions, *Auton. Agents Multi-Agent Syst.* **25** (2012), 1–45.
[21] A. M. Schmitter, 17th and 18th Century theories of emotions, in: *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.) (2010). plato.stanford.edu/archives/win2010/entries/emotions-17th18th/.

[22] M. Schröder and G. McKeown, Considering social and emotional artificial intelligence, in: *Proceedings of AISB 2010 Symposium Towards a Comprehensive Intelligence Test*, Leicester, UK, 2010.
[23] B. R. Steunebrink, M. Dastani and J.-J. C. Meyer, A formal model of emotion-based action tendency for intelligent agents, *Lect. Notes Comput. Sci.* **5816** (2009), 174–186.
[24] K. Suzuki, A. Camurri, P. Ferrentio and S. Hashimoto, Intelligent agent system for human-robot interaction through artificial emotion, in: *IEEE International Conference on Systems, Man, and Cybernetics*, 1998.
[25] W. Wallach, S. Franklin and C. Allen, A conceptual and computational model of moral decision making in human and artificial agents, *Top. Cogn. Sci.* **2** (2010), 454–485.
[26] T. Ziemke, On the role of emotion in biological and robotic autonomy, *Biosystems* **91** (2008), 401–408.