In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

EFFECT OF REWARD PREDICTION ERRORS ON THE EMOTIONAL STATE OF A MOBILE ROBOT

Vidullan Surendran (vus133@psu.edu)
The Pennsylvania State University, 234 Hammond Bldg.
University Park, PA 16802 USA

Lyle N. Long (lnl@psu.edu)
The Pennsylvania State University, 233 Hammond Bldg.
University Park, PA 16802 USA

Abstract

A goal-based dynamic action selection mechanism incorporating a model for emotions and temperament was developed for use with small and inexpensive mobile robots. A mobile robot was developed to test the action selection mechanism by recreating the scenario of an animal foraging for food while avoiding predators. Four emotions of anger, fear, happiness, and surprise were modelled which were affected by events such as finding food, encountering a predator, encountering a boundary wall, finding a safe area, and being in a state of low health. The model incorporated a reward prediction module that altered the effect of an event based on the error between when an event occurred and when it was predicted to occur. The model also included a decay term that resulted in the emotions returning to their steady state values unless there was continual reinforcement through the occurrence of events. The effect of differing temperaments on the emotions was studied by defining an irate temperament.

Keywords: Emotions; temperament; affective computing; robotics; action selection.

Introduction

Affective computing seeks to understand and develop systems that can recognize and simulate human emotions. R.W. Picard (2000) highlighted the importance of the field by exploring neurological studies that indicated human cognition was intrinsically linked with emotions. She also argues that that the development of affective computing is critical to advancing emotion and cognition theory.

Breazeal and Brooks (2005) considered cognition and emotions to be two distinct systems that evolved in intelligent creatures under social and environmental processes to aid in optimal functioning. Cognition is deemed to be responsible for interpreting the world whereas emotions are deemed to be responsible for evaluating the value of events. Emotions thus help prioritize concerns while minimizing distractions.

The simulation of human emotions is greatly complicated by the fact that there is no accepted model that explains and predicts the wide range of emotions we experience. There still is no consensus in the literature on the number of base emotions. Paul Ekman (1999) proposes a list of 15 emotions, each representing a family of emotions. A study on dynamic facial expressions of emotion by Jack et al. (2015) challenges this notion by suggesting that basic emotion communication comprises fewer categories. It is clear that our understanding of the subject is still in its infancy.

A study on momentary subjective well-being by Rutledge et al. (2014) resulted in a model of happiness referred to as the ‘Happiness Equation’ in popular culture. They showed that momentary happiness in response to a probabilistic reward is explained by the combined influence of the reward expectations and the prediction errors from the expectations; not by current task earnings as one would naively conclude. Long et al. (2015) and Long (2015) adapted this model to simulate the eight ‘universal’ emotions of fear, anger, sadness, happiness, disgust, surprise, trust, and interest in cognitive mobile robots. The emotion and temperament engine they developed was incorporated into SS-RICS which is a cognitive architecture developed at the Army Research Laboratory (Troy D. Kelley, 2006). Surendran (2015) built upon this emotional model by incorporating a reward prediction error component with a focus on developing a model that could be executed with limited computational resources. This model did not use SS-RICS and implemented a new action selection mechanism (ASM) capable of running on a small mobile robot with a Raspberry Pi processor.

Test Setup

A client-server architecture was used with an autonomous agent acting as the client, and a computer running the ASM and affective model acting as the server. A robotic platform was developed specifically for this study with an emphasis on a form factor under 400 cm^2 and reduced cost. It was based on the Raspberry Pi microprocessor and is capable of image processing at an average of 5 frames a second, orientation sensing, collision detection through infrared sensors, and battery operation for an hour.

The behavior and emotional state changes of a small rodent foraging for food while avoiding predators was chosen to be simulated. Differently colored balls having a diameter of 10 centimeters were used to represent food sources (green), predators (red), and safe areas (purple). The agent was placed inside an enclosed space containing the three types of balls which were randomly placed. One of the purple objects representing a safe area was chosen as the starting position of the robot.

The behavior that was simulated can be described as follows. The mobile robot ‘rests’ at a safe area until its health decays below a threshold level triggering a search for food; this simulates hunger. The robot then searches the surrounding area until it identifies a green ball which it tracks towards. It then 'consumes' the food by being in close proximity (within 5 centimeters) of the ball. The satiation of hunger is simulated as a time dependent increase in the robot's health for as long as it stays in close proximity to the object. When its health has been completely recharged it then seeks out a safe area to 'rest' at, until its health drops below the threshold triggering a repeat of the cycle. Throughout the simulation, the mobile robot constantly avoids red danger balls and collisions with the boundary walls. Whenever a danger is identified, the robot abandons its current task and instead takes evasive manoeuvers to avoid the danger. Only stationary predators were considered in this study. A linear decrease in health with respect to time was assumed in this study. It is to be noted that the framework allows one to implement more complex models of health and consumption.

Action Selection Mechanism (ASM)

In order for the system to be autonomous it had to decide what to do. In order for it to be successful it had to intelligently select appropriate actions based on external and internal stimuli. Dynamic planning methods compute the next action to be taken based on the current internal and external state. This type of ASM is ideal when limited computational resources are available. On the other hand, to replicate the behavior required a goal driven architecture was used (Brom & Bryson, 2015). To this end a new hybrid architecture was implemented in the final system.

The behavior implemented was decomposed into goals and subgoals. Goals were namely finding food, finding a safe area, avoiding danger and resting. These goals were then decomposed into subgoals such as finding a ball, tracking a ball, eating a ball, avoiding a ball, etc. Each of these subgoals was associated with a dynamic plan that the ASM selects in order to accomplish the goal. This modular structure permitted reuse of plans as the goals had similar subgoals. Condition-action rules similar to those used in expert systems were used to implement the dynamic plans. Rules defined the 'knowledge' the system possessed and in this system are either factual or procedural. Factual knowledge as the name suggests consists of facts such as the color of a ball defining if it relates to food, danger, or a safe area. Procedural rules govern how actions the robot take will be carried out. Conflict resolution in case of competing goals is handled using priorities with certain goals having a higher priority. For example, if a danger ball is seen in front of an objective ball, the robot would avoid the danger instead of moving towards the objective.

An event handling function and an interrupt mechanism implemented in software handle flow control. The event handler executes the appropriate plan based on the current goal, subgoal, and system state. The interrupt handler allows the currently executing goal to be paused when a goal with a higher priority is to be executed. Once complete, it reloads the previous goal and subgoal.

Time Step Each time step is defined to be one cycle of the mobile robot sending data to the command center and receiving control commands. This was chosen over time as hardware limitations prevented real time image processing and consequently governed the rate of data transmission. This ensured consistency as otherwise a slower system would experience quicker decays in health and emotions due to processing time. If enough processing power is available, the framework allows clock time to be used instead.

Emotions Engine

The computational model obtained by Rutledge et al. (2014) is shown in Equation (1). In the model CR represents certain rewards, EV their expected value and RPE the reward prediction error.

$$\begin{aligned} Happiness(t) = w_0 + w_1 \sum_{j=1}^{t} \gamma^{t-j}\, CR_j \\ + w_2 \sum_{j=1}^{t} \gamma^{t-j}\, EV_j \\ + w_3 \sum_{j=1}^{t} \gamma^{t-j}\, RPE_j \end{aligned} \tag{1}$$

Long et al. (2015) modified this model as shown in Equation (2) to incorporate emotions and temperaments into cognitive mobile robots. The winner-take-all approach they implemented meant that the emotion with the highest value was considered as the robot's emotional state. In this model R_{ij}^{+} represents positive reinforcements while R_{ij}^{-} represents negative reinforcements.

$$\begin{aligned} Emotions(t)_i = w_{0_i} \\ + \sum_{j=1}^{t} \gamma_i{}^{t-j} (w_{1_i} R_{ij}^{+} \\ + w_{2_i} R_{ij}^{-}) \end{aligned} \tag{2}$$

The emotional model used in this research study is shown in Equation (3) and incorporates the RPE term from equation (1) into equation (2).

$$\begin{aligned} Emotion(t)_i = w_{0_i} + \sum_{j=1}^{n} \gamma_i^{(t_f - t_j)} w_{1_i} R_j \\ + \sum_{j=1}^{n} \gamma_i^{(t_f - t_j)} w_{2_i} RPE_j \end{aligned} \tag{3}$$

The positive or negative effect of an event on the emotion is represented by the term R_j. The reward prediction error is represented by the term RPE. w_0, w_1, and w_2 are weighting factors. γ is a decay factor that governs the impact of past events on the current emotional state. Together these values define the temperament of an agent.

Emotions and Temperament Constants

A distinction has to be made between temperaments and emotional states. Temperaments are considered to be biologically based and derived from genetic predispositions, maturation, and experience. They are expected to be relatively stable over time. Emotions from a functionalist approach are defined as a person's readiness to establish, maintain or change one's relationship to his or her changing circumstances. In contrast to temperamental variability, emotional reactions can be enduring or brief (Thompson & Winer, 2015).

Four emotional states of anger, fear, happiness, and surprise deemed to most likely be affected by the test scenario were modelled in this study. Each emotion is assigned weighting factors and a decay factor that correspond to Equation (3) as shown in Table 1. To better illustrate the system, the weighting factors were experimentally selected to result in significant changes in emotional values within a range of ±50. The steady state value (w_0) of the emotions was set as zero for the same reason. Future work will address how to optimize these constants to make the robot as effective as possible.

Table 1: Emotion constants

Emotion	w_0	w_1	w_2	γ
Anger	0	1.9	0.15	0.92
Fear	0	1.9	0.15	0.92
Happiness	0	2.3	0.15	0.92
Surprise	0	1.7	0.15	0.88

Table 2: Emotion modifiers

	Anger	Fear	Happiness	Surprise
Danger encountered	0	+12	-5	+10
Found Food	-1	-1	+5	0
Returned to safe area	-5	-5	+5	-5
Wall encountered	+3	0	0	+5
Health too low	0	+2	-2	0

Components

The emotions engine is made up of three major components; short term memory, the RPE module, and the command modifier.

Short Term Memory The emotional state of an agent is deemed to depend on its internal state and external stimuli such as events. Events are defined as interactions with the environment that result in a change of the ASM goal or subgoal. Table 2 lists all defined events along with their reward values. The emotions engine records registered events, the reward value for the event, and the time steps taken to complete the ASM goal that led to it. Due to the decay component (γ) of the emotional model, it was found that a memory length of six events was optimal as the exponential decay meant that the value of any previous events was negligibly small.

Reward Prediction Error (RPE) Module Momentary subjective well-being was found to depend not only on an event but errors in predicting the occurrence of said event. Unexpected events have a higher impact on the value of an emotion (Rutledge et al., 2014). Due to a lack of long term memory and learning capabilities, the average number of time steps taken for an event to occur in the past is used to predict future expectations. The difference between this prediction and the actual time taken for the event to occur is considered to be the RPE. At the start of an agent's life each type of events is predicted to take 25 time steps to occur. An exception is the 'Health too low' event which is considered to take 300 time steps to occur. These values were chosen as they were found to be the average number of time steps taken in the majority of tests. As events of a type occur during operation, the prediction for that event is updated.

Command Modifier Data stored by the short term memory module is sent to the RPE module to calculate the RPE. Using Equation (3) the instantaneous value of the four emotions are then calculated and in combination define the emotional state of the robot. These calculations are carried out every time step regardless of an event being registered. The command modifier allows the emotions engine to modify the ASM commands generated and the internal state of the agent, based on the current emotional state. Conditional logic and statements are used for this purpose. In this study we have chosen to apply the following modifiers,

- $Speed = current\ speed + value_a + value_f - value_h$

Where $value_a$, $value_f$, and $value_h$ denote the numeric values of the emotions anger, fear, and happiness respectively. This modifier increases the speed of the robot when the value of anger or fear increases and decreases it when happiness increases. This shows how all emotions can compete against each other to affect behavior.

- $if\ speed < 90\ then\ speed = 90$

This modifier sets a lower bound on the speed of the robot.

- $if\ value_f > 40\ then\ randomly\ stop\ the\ robot's\ motion$

This simulates a timid robot by randomly sending it stop commands if the value of fear becomes greater than 40. The is similar to the tendency of an animal to momentarily freeze when frightened.

- $if\ surprise > 40\ then\ make\ a\ 530°\ turn$

This simulates a surprised robot by making it turn around if the value of surprise gets larger than 40. This is similar to an animal being startled and a robot spinning around was not only amusing, but easily observable.

Results

The effects of the different type of events, the RPE module, and varying temperaments were considered. The test scenario pictured in Figure 1 consisted of two safe balls, three food balls, and three danger balls. It was run for 300 time steps.

Figure 1: Test Scenario

Events are signified on the plots using the letters shown in Table 3. A negative RPE denotes an unexpected event while a positive RPE denotes an event that was predicted to occur earlier than observed. Table 4 lists all events that were registered along with their time steps and the calculated RPE.

Table 3: Events and their signifiers

Event	Denoted by
Danger encountered	a
Found Food	f
Returned to safe area	b
Wall encountered	w
Health too low	h

Table 4 Time steps, Event type, and RPE value of all events

Step	**22**	**23**	**24**	**35**	**43**	**51**	**59**
Event	f	f	f	w	w	w	w
RPE	-13	-5	-1	10	-22	-11	-5
Step	**85**	**92**	**107**	**146**	**158**	**181**	**204**
Event	a	w	w	a	w	w	w
RPE	60	23	-6	6	33	-11	-4
Step	**214**	**217**	**225**	**231**	**241**	**242**	**248**
Event	b	w	w	w	f	w	f
RPE	164	-14	-11	-7	12	2	13
Step	**249**	**255**	**256**	**260**	**266**	**294**	
Event	w	f	w	f	f	w	
RPE	-3	14	-1	12	12	31	

Effect of Events on the Emotional State

Figure 1 shows the variation in the individual emotions whereas Figure 2 shows the variation in the robot's speed and its health. From the data we can infer that, once the health dropped below the threshold of 75 at time step 10, the agent began searching for food and found it at time step 22, that is in 12 steps. The RPE led to a larger spike in happiness and a large reduction in robot speed. During its search for a safe area it had to avoid the boundary walls, food balls, and the danger balls each of which affected the emotions as seen in the previous scenarios. The robot finally found a safe area at time step 214. The initial prediction by the RPE module was that a safe area would be reached in 25 time steps. In this case since that led to an RPE of 164, instead of an increase in happiness and reduction in anger, surprise, and fear, we see the opposite effect. Since the robot took a large number of time steps to return to a safe area, when it did its health had dropped below the threshold of 75. This caused it to immediately start searching for a food ball. At approximately time step 241 there was an anomaly where the robot incorrectly identified a food ball as a boundary wall momentarily before correctly recognizing it as a food ball.

Effect of the RPE Module

Figure 3 shows the effect of the RPE on the emotional state value of anger. Using Table 4 we study some key events that illustrate the effect of the RPE module:

Food event at time step 22 (RPE of -13) This event happened sooner than predicted by 13 time steps. The effect of the RPE module was a greater reduction in anger than would have been without it.

Wall event at time step 35 (RPE of +10) This event happened later than predicted by 10 time steps and thus its impact was reduced. This is seen as anger values not being as large as they would have without the RPE module.

Wall event at time step 43, 51, 59 (RPE of -22, -11, -5) We observe that the impact of the wall at step 43 is greater than that at step 51 and so on. When the RPE module was disabled, the impact of all 3 wall events was exactly the same.

Safe area ball at time step 214 (RPE of 164) The extremely large prediction error meant that instead of a decrease in anger that would have been expected when the robot returned to a safe area, there was an increase in anger. Basically the opposite of the effect the event would have had on the emotions had the RPE module been disabled.

We can thus conclude that the RPE module is crucial in modelling the psychological effect of expectations governing the effect of an event. Without it, an event would affect the emotional state exactly the same way regardless of when it occurs.

Effect of Varying Temperaments

Temperaments are defined by the weighting factors shown in Table 1.

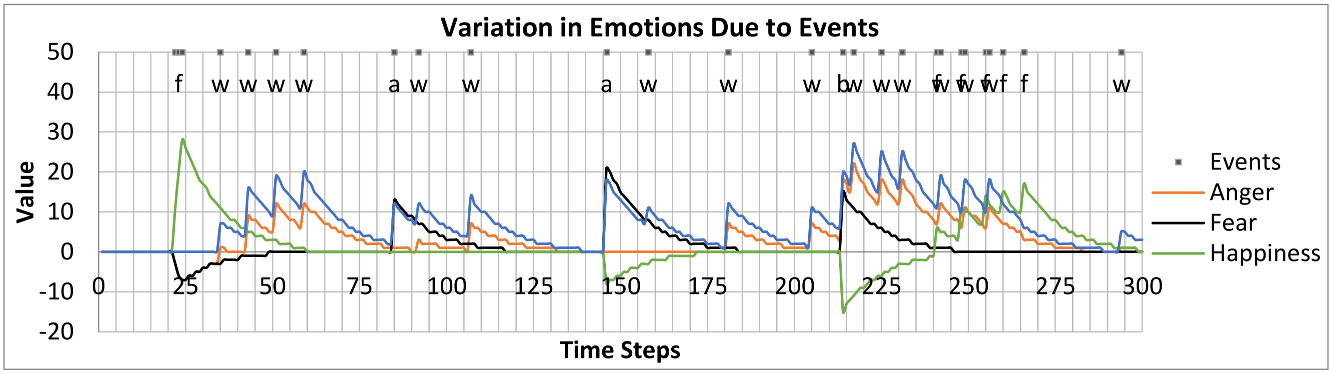

Figure 1: Shows the variation in the values of the four emotions

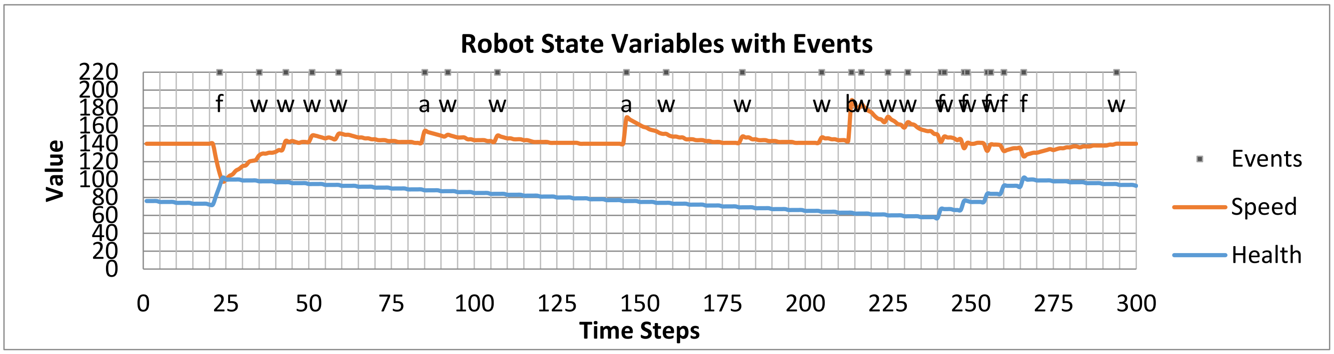

Figure 2: Shows the variation in the speed and health of the robot for a test scenario

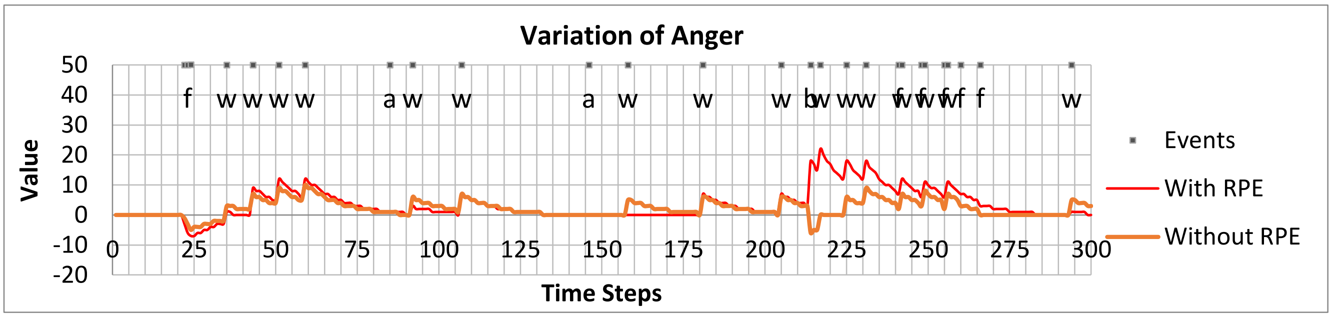

Figure 3: Compares the variation of anger with and without the RPE module

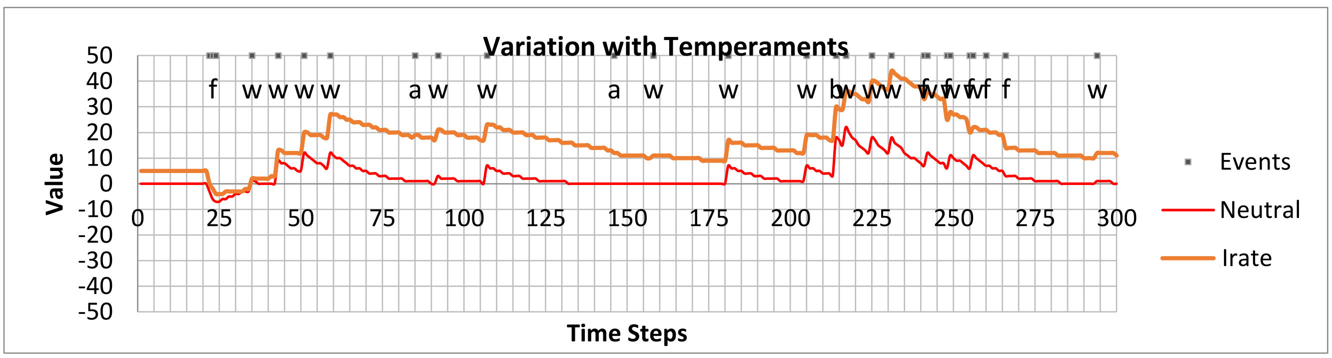

Figure 4: Compares the value of anger for an irate and a neutral temperament

Figure 4 shows the change in the value of the emotion anger for an irate temperament which was defined by modifying the weighting factors affecting the emotion anger.

Irate: The weighting factors used for anger were:

Emotion	w_0	w_1	w_2	γ
Anger	5	2.3	0.15	0.98

We observe that the steady state value of the affected emotion is offset from 0 by the value of the weighting factor w_0. The factor w_1 controls the instantaneous effect of an event on the emotions. As γ is increased we see that an emotion takes longer to decay to its steady state value. Thus an irate temperament causes the robot to have a higher steady state value of anger and remain angry for longer once an event angers it.

Conclusions

By adapting a model for momentary well-being (Rutledge et al., 2014) as done by Long et al. (2015), and incorporating a reward prediction error, an action selection mechanism with an integrated emotions engine has been implemented in addition to the development of a low cost robot to test the mechanism (Surendan, 2015). The ASM developed was a hybrid of a goal-based and a dynamic planning action selection mechanism. Each goal was associated with subgoals, and each subgoal with a dynamic plan. Different events were defined that affected the individual emotions and the ASM provided means by which the emotional state could be used to modify the internal state of the robot through means of command modifiers. Command modifiers also allowed the modification of dynamic plans and the value of emotions to be coupled through suitable models. Temperaments and emotional variability were defined using a matrix of constants. Varying the temperaments was observed to result in a different emotional state over the time period of the experiment even when the scenario was kept constant. Finally, the importance of the reward prediction error was highlighted by showing that without it events affected the emotions by the same amount regardless of when they occurred. With the RPE, it was possible to mimic the emotional response observed in humans by Rutledge et al (2014).

Additional tests should be conducted by specifying a more comprehensive list of command modifiers and fine tuning the temperament values based on the observation of an animal foraging for food in the wild. Since the temperament is specified by means of constants shown in Table 1, this leads to the possibility of an agent with a time-variant temperament that can alter its emotional sensitivity during run time. The emotional model could also be modified to introduce a time lag between the occurrence of an event and it affecting the emotional state. This would allow the simulation of the four classic temperaments, namely, melancholic, phlegmatic, choleric, and sanguine theorized by Greek philosophers ("Four Humors - And there's the humor of it: Shakespeare and the four humors," 2012). Another approach would be the specification of the temperament in terms of the "Big 5" traits of extraversion, agreeableness, openness, conscientiousness, and neuroticism often described in the literature (Digman, 1990). The low cost of the robot also permits the acquisition of a non-homogeneous robot swarm with varying temperaments to explore if emotions increase the effectiveness of the swarm.

References

Brom, C., & Bryson, J. (2006). *Action selection for Intelligent Systems*. European Network for the Advancement of Artificial Cognitive Systems.

Breazeal, C. & Brooks, R. (2004). Robot Emotion: A Functional Perspective. In *Who Needs Emotions,* J. Fellous (Ed.), Oxford University Press.

Digman, J. M. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, *41*(1), 417–440.

Ekman, P. (1999). Basic emotions. In *Handbook of cognition and emotion*. Wiley & Sons Ltd.

Four Humors - And there's the humor of it: Shakespeare and the four humors. (2012, January 30). Retrieved from https://www.nlm.nih.gov/exhibition/shakespeare/fourhumors.html.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2015). Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. *Current Biology*, *24*(2), 187–192.

Long, L. N. (2015). Modeling Emotion and Temperament on Cognitive Mobile Robots. *Presented at the 22nd Annual ACT-R Workshop*. Pittsburgh PA: Carnegie Mellon University.

Long, L. N., Kelley, T. D., & Avery, E. S. (2015). An Emotion and Temperament Model for Cognitive Mobile Robots. *Presented at the 24th Conference on Behavior Representation in Modeling and Simulation (BRIMS)*, Washington, DC.

Thompson, R. A., & Winer, A. C. (2015). The individual child: temperament, emotion, self, and personality. In *Developmental science: An advanced textbook, Edition: 6*. Psychology Press / Taylor & Francis.

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257.

Picard, R. W. (2000). *Affective Computing.* Cambridge, MA: The MIT Press.

Kelley, T. D. (2006). Developing a psychologically inspired cognitive architecture for robotic control: The symbolic and subsymbolic robotic intelligence control system (SS-RICS). *Advanced Robotic Systems*, *3*(3), 219–222.

Surendran, V. (2015). *A dynamic action selection mechanism for an autonomous agent with a model for its emotional state*. M.S. thesis, Aerospace Engineering, The Pennsylvania State University, State College, PA.