In Proceedings of the Sixth International Conference on Cognitive Modeling, 172-177. Mahwah, NJ: Lawrence Earlbaum.

Toward a Comprehensive Computational Model of Emotions and Feelings

Robert P. Marinier III (rmarinie@umich.edu)
University of Michigan, 1101 Beal Avenue
Ann Arbor, MI 48109-2110 USA

John E. Laird (laird@umich.edu)
University of Michigan, 1101 Beal Avenue
Ann Arbor, MI 48109-2110 USA

Abstract

We describe a computational framework for emotions and feelings that combines biological, cognitive, and social influences. The framework, Soar-Emote, combines ideas from Damasio and Gratch & Marsella. From Damasio we include physiological influences, ideas about the differences between emotions and feelings, and the direction of causality. From Gratch we include cognitive and social influences, appraisal theory and coping. We give results that show that these systems influence agent behavior for a simple interactive game.

Introduction

If one thing is clear in the field of computational models of affective reasoning, it is that the exact phenomena and behavior that falls under the rubric of "emotion" or "affective reasoning" is very murky. Some theories of emotion attempt to model low-level physiological effects of specific emotions in isolation, whereas others attempt to model the interaction between of emotion and cognition, while others attempt to model the impact of emotional responses on social interactions, typically without attempting to model any of the underlying physiological processes. An analysis by Biddle et al. (2003) of a sampling of current computational models showed that several systems attempt to model affective behavior in one or two of these areas, but no systems attempts to model behavior across all. Our hypothesis is that affective reasoning is not an isolated phenomenon and that to understand it we need to build computational models that have impact across all of these areas. Further, such models must be intertwined with a theory of cognition – emotion is not a separate "chip" (as often portrayed in science fiction AI systems) but is integral to cognitive processing.

This paper describes a computational framework for affective reasoning. We chose two theories that together cover biological, cognitive and social influences on emotions and feelings. Damasio (1994, 2003) describes the role of the body in emotions and feelings as well as the high-level architecture of the human affective system. Damasio also makes an important distinction between emotions and feelings, but only touches on the role that cognition plays in processing and responding to emotions. Furthermore, Damasio's models are descriptive, without the accompanying precision of computational implementations. Gratch & Marsella (2004) on the other hand, focus on a computational model of the interaction of affect with cognition and the social level, including how feelings are represented in a general cognitive architecture, Soar (Newell 1990), and how coping is used to respond to emotions.

Our framework refines a subset of the Gratch model, adding components to model the concepts of Damasio. Moreover, our model, which is an instantiation of the framework, involves *architectural changes* to Soar, so that affective reasoning is directly incorporated into Soar, in contrast to the Gratch implementation where affective reasoning was implemented as knowledge (encoded as rules) in the cognitive system. The research is not mature enough to make precise predictions concerning the generation of specific emotions and feelings in humans. Instead, our goal is to demonstrate that the framework captures important interactions between physiology, cognition, and social interaction. We do this by embedding the model in agents that play a simple cooperative game where emotions can arise. We introduce "lesions" in the agents that disable different aspects of the framework, and confirm that each aspect of our framework contributes to the system's overall behavior.

Prior Work

Gratch's work is notable because it is a model of how emotions and feelings impact cognition and how cognition responds to feelings. Gratch's system, called EMA, uses appraisal theory[1] to describe how an agent evaluates a situation, and how the resulting feeling is represented to the agent. Thus, emotional responses are not hardwired into inputs from the perceptual system, but instead are based on the agent's interpretation (appraisal) of the current situation, which can include arbitrary cognitive processing or even the recall of prior situations from memory. He also describes various coping mechanisms that involve cognitive processing. These mechanisms are strategies (once again, not hardwired responses) that the agent can selectively employ to respond to feelings and include both problem-focused coping strategies, in which an agent takes action in

[1]Appraisal theory is widely studied and many versions exist; for a survey, see Ellsworth & Scherer (2003). In this work we are primarily concerned with the version in EMA.

the world, and emotion-focused strategies, in which an agent reevaluates a situation cognitively.

EMA consists of two basic steps that cycle repeatedly. The agent appraises the current situation, which results in an "emotion." The agent then copes with the "emotion," either by altering its appraisals (emotion-focused) or by altering its situation (problem-focused).

EMA defines several appraisal variables, including Desirability, Likelihood, and Attribution. Desirability is a measure of how much the agent wants to be in the situation. Likelihood, which is the probability of a situation being true, is useful when an agent is considering situations that may have been or might be (either now or in the future). Attribution is the thing that caused the situation, although there may not always be an object for attribution. Combinations of these variables in different ranges give rise to different feelings, as shown in Table 1. Furthermore, the intensity of the feeling is the absolute value of the Desirability multiplied by the Likelihood.

Table 1: "Emotions" (feelings) defined by EMA.

"Emotion" (Feeling)	Aggregated Appraisals
Joy	Desirability > 0 Likelihood = 1
Hope	Desirability > 0 Likelihood < 1
Fear	Desirability < 0 Likelihood < 1
Dismay	Desirability < 0 Likelihood = 1
Anger	Desirability < 0 Object

An example of an appraisal might involve the agent getting dirty because another agent spilled coffee on it. The situation gets a negative desirability, and the likelihood is 1.0 since it actually happened. If the agent views it as an accident, the feeling will come out as dismay. However, if he blames the other agent, he may feel angry. In fact, if he merely suspects that the other agent did it on purpose, then he may feel fear.

A problem-focused coping strategy may involve getting cleaned up. An emotion-focused coping strategy reappraises the situation in a different light. For example, an emotion-focused coping in this situation may involve thoughts along the lines of, "It was going to get dirty eventually anyway."

EMA has some drawbacks for modeling emotions and feelings. Its model of emotion is implemented as knowledge about how to feel in light of various appraisals instead of having an *independent* emotion system that *interacts* with knowledge processing. Furthermore, EMA does not incorporate any notion of the non-cognitive effects the body may have on the agent's affective state. Even if the system were to include the physiological state of the body, this would only influence the agent's affective state through cognitive appraisal.

Damasio, on the other hand, describes the role of the body in determining affect. He also provides the distinction between emotions and feelings, as well as a high-level description of how bodily notions of emotions and feelings interact with cognition. Figure 1 gives a high-level overview. An emotionally competent stimulus affects the body. For our purposes, an emotionally competent stimulus is a stimulus that impacts the emotions of the agent. This stimulus is appraised, both cognitively and bodily. This appraisal causes a change in the state of the body (i.e. the physiology) – this is what Damasio calls the emotion. It is the perception of this emotion that is the feeling. This feeling impacts the decisions made by the agent, which then leads to actions that can affect the environment (and hence the perceived stimulus). Damasio also reports that humans can have non-cognitive emotional responses, even when they are unaware of the emotionally competent stimulus.

There is not complete agreement between Damasio's model and Gratch's. First, there is a vocabulary difference. Gratch uses emotions and feelings interchangeably, whereas Damasio provides a clear distinction. In particular, Damasio distinguishes between the physical state of the body (the emotion) and the perception of that emotion (the feeling). This distinction is not useful for Gratch since EMA does not include a model of physiology. Even so, what Gratch calls an "emotion", Damasio would consider feeling. EMA deals with cognitive interaction, and in order to do this the "emotion" must be represented in working memory – that is, the agent must be aware of it. This awareness is a perception of the emotion, and hence, by Damasio's terminology (which we adopt), it is a feeling. Furthermore, in EMA, feelings come first, and then the emotions. That is, once an agent determines that it feels bad, it generates the appropriate body language to display that – Damasio claims the order is actually reversed, based on several brain imaging and case studies he reviews. Additionally, EMA generates feelings via rule firings that are part of the cognitive system and could be open to the influence of other knowledge as well as cognitive learning, as opposed to a more fixed and architectural mechanism involving the body as suggested by Damasio.

Our framework combines these models to produce one that takes into account both physiological and cognitive effects on emotions and feelings. We use Damasio's theory regarding the overall system architecture and physiological

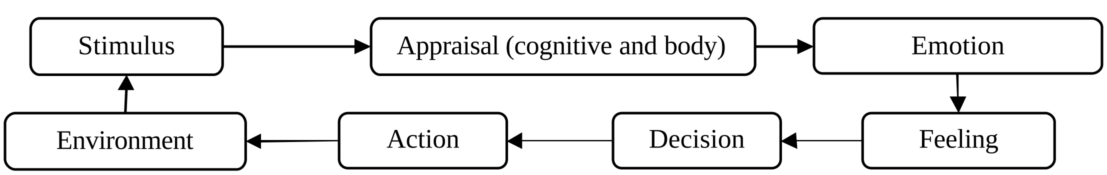

Figure 1: Basic order of events in Damasio's model.

details, but rely on Gratch's approach for modeling appraisal and coping. The challenge is to determine how the body and cognition interact to produce emotions and feelings. Furthermore, we need to reverse the direction of causality in EMA for feelings and physiological effects.

A Simple Cooperative Game

To demonstrate our framework, we have implemented a simple simulated water balloon toss game between two players. In this game, two players toss water balloons back and forth, attempting to score points by successfully catching the balloons (this is cooperative and not competitive); however, things can go awry so that a balloon is missed and one player gets wet or one player might have to dive into bushes (and experience pain) to catch the water balloon. These interactions with the environment and the agents' perception of the body state of each other can give rise to emotions and feelings that impact the game play and lead the agents to remark about each other's performance.

In the game, the players stand in a field of grass and bushes (this is not a full 3D simulation, but a simple abstract simulation of these two possible situations), with one of them the thrower and the other the catcher. The thrower tosses the balloon to the catcher, who then tries to catch it. Following the attempted catch, the thrower gets to remark on the situation. Once the thrower has remarked, the catcher gets to remark. The thrower then gets to consider the catcher's remark, but does not get to take action. Then the next round begins, with the thrower now being the catcher and vice versa.

When the thrower tosses the balloon, he can vary two parameters: the speed (slow or fast) and the aim (near the catcher or far away). Similarly, the catcher can vary two parameters: whether he runs (true or false) and whether he makes an attempt to catch the balloon (true or false). A player can only move when he is the catcher, and then only to the balloon's landing location. There are also two parameters for the remarks: subject (you or me) and type (supportive or critical). A player can also choose silence.

While the thrower may intend to throw a particular way, the result may come out differently (i.e. he may throw fast and far when he meant to throw slow and near). The catcher only sees the actual throw. If the balloon is caught, the players get a point. Otherwise, the catcher may get wet. They have an infinite supply of balloons, so the game can continue even if the balloon is not caught.

Soar-Emote

Our framework has a simplified version of the basic appraisal model used in EMA, but instead of a rule-based implementation of the emotion generation, we introduce architectural mechanisms that are fixed and are used for all tasks. The main differences on the cognitive side consist of an architectural mechanism called the Appraisal Summarizer that translates the appraisals into the cognitive contribution that is then sent to the body. This cognitive contribution is not an emotion, but instead represents the influence of cognition on the emotion. At the same time, the body's Internal Physiology creates an appraisal of the current body state. The Emotion System combines this information from Internal Physiology and the Appraisal Summarizer to create the actual emotion of the system. The agent can have non-cognitive responses to the emotion, such as automatic changes to body language. The perception of this emotion is the feeling the agent gets, which it can then cope with, as per EMA. Like Gratch, we use Soar for the basis for the cognitive system.

Figure 2 shows the overall architecture of a single agent with the cognitive system being on right, and the physical (i.e. non-cognitive) system on the left. The remainder of this section describes the components of the framework and their interactions within the context of the water balloon toss game as told from the point of view of the catcher. Each component is numbered according the descriptions below. We assume that just prior to the beginning of the example, the thrower tossed the balloon far and fast, and the catcher ran and tried to get it and missed. Then the thrower made a critical remark about the catcher. It is now the catcher's turn to remark. There are no specific critical remarks in the program – the agents can just detect whether a remark is critical or not; however, to bring the example to life, you might imagine that the thrower just said, "Nice try Crisco hands!"

(1) External stimuli from the environment arrive at the agent's body. Some stimuli are picked up by perception. In this example, the agent sees that the thrower looks angry, and also hears a critical remark about himself.

(2) External stimuli can also directly affect the agent's physiology, which in turn can be perceived. For example, changes in body temperature that result from the external temperature can be perceived. In this example, the agent is not feeling pain but he is tired since he just ran.

(3) Perception processes stimuli from both external and internal sources and sends them to other systems, namely Working Memory and possibly the Motor System (in the case of reflexes).

(4) In some cases, a stimulus may lead directly to the creation of a motor command. For this simulation, however, reflexive actions were disabled.

(5) Percepts from Perception enter Working Memory. These percepts include processed forms of external stimuli and internal stimuli. Internal stimuli include perceptions of the internal physiology and the current emotion (the perception of which is a feeling). In this example, the current feeling is one of dismay (this stemmed from the immediately prior situation in which there was a bad throw). The agent is also tired and hot but not in pain. The agent becomes aware of these and other details about the situation.

(6) Working Memory contains the objects the agent is currently "thinking" about. These include objects that originated external from the agent's mind (i.e. from the body or the environment) as well as objects it has created internally (i.e. objects it uses in the process of solving problems or otherwise behaving).

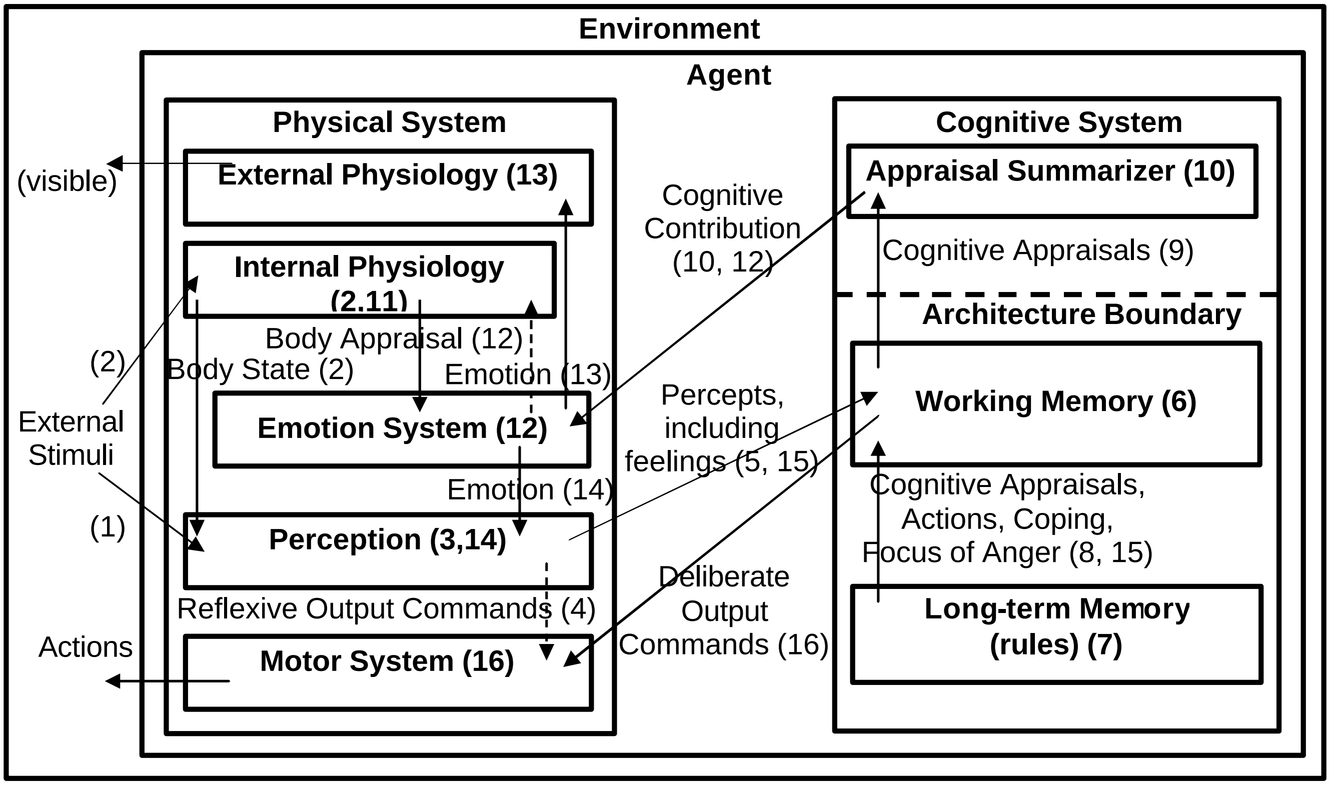

Figure 2: Overview of the Soar-Emote model.

(7) Long-term memory contains rules that encode the agent's knowledge. Thus, objects in working memory may trigger a rule. The kinds of rules present include rules about how to perform actions, such as throwing the balloon and how to evaluate a current situation. Rules to evaluate the situation may be triggered by numerous emotionally competent stimuli; in this case, the evaluations include the current score, the failed catch, the other agent's body language, and the remark.

(8) When rules match and fire, they make changes to Working Memory. In this example, rules fire that create negative appraisals for the agent's body temperature, exertion, failure to catch the balloon, the other agent's angry body language, and the negative remark. In fact, the body language contributes to two appraisals: it gets an appraisal in and of itself, and it also provides context for the remark. Thus, the agent realizes that this is not constructive criticism. Furthermore, the agent attributes some of these, such as the failure to catch the balloon and the negative remark, to the other player.

(9) Appraisals are automatically transferred into the Appraisal Summarizer, which in this case is an extension of the Soar architecture.

(10) The Appraisal Summarizer summarizes these cognitive appraisals and sends the result, which is the cognitive contribution to the next emotion. The intensity of a cognitive contribution is the product of the average desirability and average likelihood. This approach is inspired by EMA, although there are several differences. In EMA, this summary is the next "emotion," whereas in Soar-Emote it is just a contribution. Also, in Soar-Emote, all of the numeric appraisal variables of a particular type are averaged together, and the resulting values are used to generate the cognitive contribution. Instead of this averaging process to produce a single "emotion," EMA allows degrees of multiple emotions at once. The primary reason for this difference at this point is simplicity. The purpose of the framework at this stage is to show how the physical and cognitive systems interact to alter feelings and behavior; thus, the full complexity of EMA is unnecessary.

In this example, the overall desirability is negative and the intensity of appraisals with attributions outweighs those without attributions, so anger is generated for the cognitive contribution.

(11) While all of this activity in the Cognitive System is going on, Internal Physiology generates an appraisal of the current body state based on the agent's body temperature, exertion, pain, and noise. In this example, the agent's temperature is high and he is tired, but he is not in pain. Thus, the body appraisal is negative; however in this case we assume that the added noise is positive, lessening the negative body appraisal.

(12) The contribution from the Cognitive System and the body appraisal from Internal Physiology are combined by the Emotion System into an emotion that the agent will experience. The generation is biased towards the cognitive contribution. In this example, the agent's emotion is still one of anger, but combination with the body appraisal decreases its intensity.

(13) The emotion generated from the Emotion System is sent to External Physiology, which determines the agent's body language. Body language is an abstraction that encompasses externally perceivable aspects of the agent's emotion, such as facial expression and tone of voice. In the current framework, the agent does not have any direct control over his own body language. In our example, the agent is feeling angry. However, it is not very intense, so the body language becomes neutral.

(14) The emotion generated from the body's emotion system also feeds into perception where it becomes the agent's feeling (recall that feelings are the agent's perception of its emotions).

(15) The feeling, along with the latest percepts, are transferred into Working Memory. The agent can appraise the situation again and it can cope with its feelings. It can also explore why it is feeling a certain way; for example, to determine the primary source of anger. All of these are done via the application of knowledge from Long-Term Memory. In Soar-Emote, the agent may reappraise a situation repeatedly until his feeling stabilizes.

In this example, the agent first engages in emotion-focused coping by positively reinterpreting the fact that he failed to catch the balloon. Thus, the cycle of generating a new feeling must repeat. While this reduces the intensity of his anger, he is still angry. He identifies the thrower as the source of his anger, and then he decides to use problem-focused coping by making a critical remark about the thrower (in an attempt, perhaps, to let the thrower know he is angry with the situation).

(16) Output commands are sent to the Motor System, which generates actions in the world, which in this case is consists of a negative remark being made by the agent.

Table 2: Systems involved at each level.

Level	Systems
Biological	Internal and External Physiology, Emotion System
Cognitive	Appraisal Rules, Appraisal Summarizer, Emotion-focused coping
Social	Problem-focused coping (remarks), Perception of External Physiology of others

Influences at Each Level

It is easy to lose the forest for the trees in the previous description of Soar-Emote. Recall that one of the goals for our framework is to be comprehensive and cover the impact of emotion and feelings at three levels: biological, cognitive, and social. Table 2 summarizes how the framework influences these areas.

Results

Our aim is to answer the questions: Do each of the high-level aspects of the system (biological, cognitive, social) contribute to the agent's emotions and feelings enough to change its behavior? What are these differences? It is premature to compare the model's performance to human data; however, we can examine whether the framework achieves our goal of incorporating influences from each of the levels. The agent has a default action for each phase that does not require a feeling (i.e. non-emotional behaviors); for throwing this is the near/slow throw, for catching it is the run/attempt catch, and for remarking it is silence. Other action choices are motivated by emotions and feelings. To demonstrate that the components of Soar-Emote influences behavior across all levels (biological, cognitive, social), we created five types of agents with various "lesions", where groups of components were disabled. Table 2 shows that components correspond to each of the levels we modeled; the non versions of these agents had most of these components disabled. The exception was the Emotion System – for the Non-Biological agent, the agent still experienced feelings, but they were based entirely on the cognitive contribution. Additionally, we tested an Affective agent that had all systems enabled, and a Non-Affective agent that had all of these systems disabled. The Non-Affective agent demonstrates what a Soar agent without any of these augmentations would do. The test consisted of pairing each agent with an agent of the same type for 100 games of 20 rounds each.

In general, agents with different combinations of active components showed different variations in behaviors, stemming both directly and indirectly from differences in the emotions and feelings the agents experienced. We show the results for the remarking and catching behaviors because they had the most variation. Throwing behaviors showed much less variance primarily because the agents do not remember history between rounds and emotions and feelings in our system do not currently have any momentum (they exist only as long as the stimuli that cause them).

The Non-Biological agent does the run/attempt catching behavior significantly more often than the fully-affective agent, and it never choose the attempt only catching action (Figure 4). This suggests that the Non-Biological agent feels better than the fully affective agent most of the time, since not trying the run/attempt catch is indicative of negative feelings (for example, the agent may be angry and refuse to catch the balloon).

The Non-Cognitive agent chooses the silence remark significantly less than the fully affective agent (Figure 3). It also chooses the critical/me remark much more, but never chooses critical/you. This is because without the cognitive influence, the agent has no way of assigning blame, and hence never feels angry at others. However, he can feel dismay, and hence remark critically about himself.

The Non-Social agent always chooses silence, which is a significant departure from the fully affective agent (Figure 3). The Non-Social agent also always chooses the default throw (not shown). This is because, in the extra time between the catch and the next throw, nothing happens, that gives the emotions time to settle down.

In general, the throws (not shown) vary little. This is because the agent lacks history – the remarks happen at the end of the round, and so are immediately forgotten when the

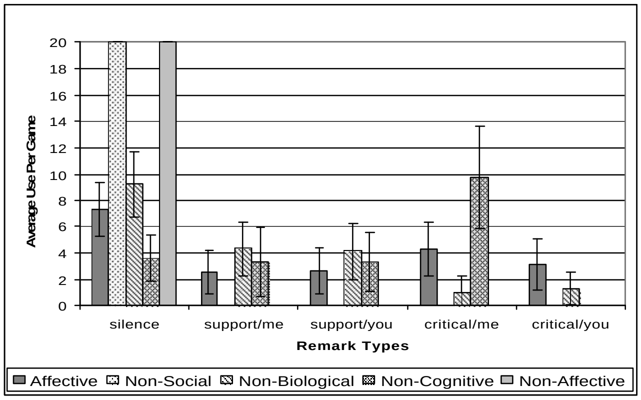

Figure 3: Remarking behaviors by various agent types.

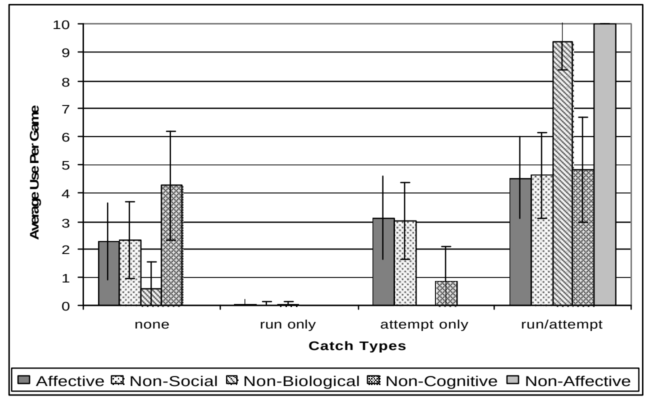

Figure 4: Catching behaviors by various agent types.

next round starts and hence have very limited impact. To rectify this, we added basic history capabilities to the model, and the variability in the throwing behaviors increased significantly, especially for the Non-Social agent.

Overall, we see that the biological parts of the system has a fairly negative impact on the emotions and feelings, whereas the cognitive parts has a more positive impact which helped mediate the biological impact. In general this is probably not true in humans, but given the task and the parts of the agent we are modeling, it is plausible. The social impact was primarily seen in the remarking. The reason the social impact was so limited everywhere else is partially because of the lack of history – when we added history, the non-social agent experienced the most change in the throwing round. The other issue is that the agent's social abilities are very limited, such as a lack of knowledge about what is socially acceptable behavior.

Conclusion

Our primary goal has been to strive towards a comprehensive cognitive framework for emotions and feelings by including biological, cognitive and social influences. We modified the Soar architecture to include ideas from Gratch, and then unified that system with biological systems inspired by Damasio. The full framework includes direct biological and cognitive influences on emotions and feelings. Social influences come in the form decisions to make various remarks, and the appraisals of those remarks and the body language of other agents, and our results show that the influences of these areas had an impact on the agents' behavior.

While we do address several issues, there are many phenomena that the framework does not address adequately or at all. Soar-Emote is a starting point we plan to build on and expand. Future work with Soar-Emote's includes work at all of the levels. For example, biologically, Soar-Emote lacks emotional momentum and non-cognitive modification of the perception of emotionally competent stimuli (i.e. as in the experience of sympathy). Cognitively, Soar-Emote lacks the ability to cognitively moderate automatic emotional responses (Fellous, 2004). We also need to explore the impact of emotions and feelings on the cognitive architecture's processing, such as influencing long-term memory retrieval (Hudlicka, 1997). Socially, our results indicate that history will play an important role. Other changes that impact multiple levels include a more complete theory of appraisal (Smith, 2004), and an explanation of individual differences that goes beyond differences in long-term knowledge (Hudlicka, 1997).

We expect that changes to the framework to incorporate these and other phenomena will involve the creation of new and more complex subsystems, modification of the Soar architecture, and new kinds of rules to take proper advantage of these changes.

References

Biddle, E. S., Henninger, A., Franceschini, R. & Jones, R. M. (2003). Emotion Modeling to Enhance Behavior Representation: A Survey of Approaches. *Interservice/Industry Training, Simulation, and Education Conference*.

Damasio, A. (1994). Descartes' Error: Emotion, Reason and the Human Brain. New York, NY: Avon Books.

Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Orlando, FL: Harcourt, Inc.

Ellsworth, P. & Scherer, K. (2003). Appraisal Processes in Emotion. In Davidson, R., Scherer, K. & Goldsmith, H (Eds.), *Handbook of Affective Sciences*. Oxford: Oxford University Press.

Fellous, J. (2004). From Human Emotions to Robot Emotions. *AAAI 2004 Spring Symposium on Architectures for Modeling Emotion: Cross-Disciplinary Foundations.*

Gratch, J. & Marsella, S. (to appear). A domain-independent framework for modeling emotion. *The Journal of Cognitive Systems Research*.

Hudlicka, E. (1997). *Modeling Behavior Moderators In Military Human Performance Models*. Technical Report 9716. Arlington, MA: Psychometrix Associates, Inc.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Smith, C. A. A (2004). Functional Perspective on Emotion Elicitation. *AAAI 2004 Spring Symposium on Architectures for Modeling Emotion: Cross-Disciplinary Foundations.*