Appraisal in Human-Robot Collaboration

Mahni Shayganfar, Charles Rich, Candace L. Sidner
Worcester Polytechnic Institute
Comoputer Science Department
100 Institute Road
Worcester, Massachusetts 01609
mshayganfar | rich | sidner @wpi.edu

Abstract

Don't know yet!

Introduction

- SharedPlans theory.
- Appraisal theory and social context.
- The necessity for identifying underlying processes of the collaboration.

Example Scenario

Two short interactions of the robot and the astronaut (to be used throughout the paper).

Affective Motivational Collaboration Theory

The Affective Motivational Collaboration Theory is built on the foundations of the *SharedPlans* theory of collaboration (Grosz and Sidner 1990) and the *cognitive appraisal* theory of emotions (Gratch and C.Marsella 2004). Affective Motivational Collaboration Theory is about the interpretation and prediction of observable behaviors in a dyadic collaborative interaction. The theory focuses on the processes regulated by emotional states. The observable behaviors represent the outcome of reactive and deliberative processes related to the interpretation of the self's relationship to the collaborative environment. Affective Motivational Collaboration Theory aims to explain both rapid emotional reactions to events as well as slower, more deliberative responses. The reactive and deliberative processes are triggered by two types of events: *external* events, such as the other's *utterances* and *primitive actions*, and *internal* events, comprising changes in the self's mental states, such as belief formation and emotional changes. Affective Motivational Collaboration Theory explains how emotions regulate the underlying processes when these events occur during collaboration. This theory elucidates the role of motives as goal-driven emotion-regulated constructs with which an agent can form new intentions to cope with internal and external events (cite my own AAAI paper).

Affective Motivational Collaboration Theory explains the functions of emotions in a dyadic collaboration and show how affective mechanisms can coordinate social interactions by enabling one to anticipate other's emotions, beliefs and intentions. Our focus is on the mechanisms depicted as mental processes in Figure 1 along with the mental states. The *Mental States* includes self's (robot's) beliefs, intentions, motives, goals and emotion instances as well as the anticipated Mental States of the other (human). The *Collaboration* mechanism maintains constraints on actions, including task states and the ordering of tasks. The *Collaboration* mechanism also provides processes to update and monitor the shared plan. The *Appraisal* mechanism is responsible for evaluating changes in the self's Mental States, the anticipated Mental States of the other, and the state of the collaboration environment. The *Coping* mechanism provides the self with different coping strategies associated with changes in the self's mental states with respect to the state of the collaboration. The *Motivation* mechanism operates whenever the self a) requires a new motive to overcome an internal impasse in an ongoing task, or b) wants to provide an external motive to the other when the other faces a problem in a task. The *Theory of Mind* mechanism is the mechanism that infers a model of the other's anticipated mental state. The self progressively updates this model during the collaboration.

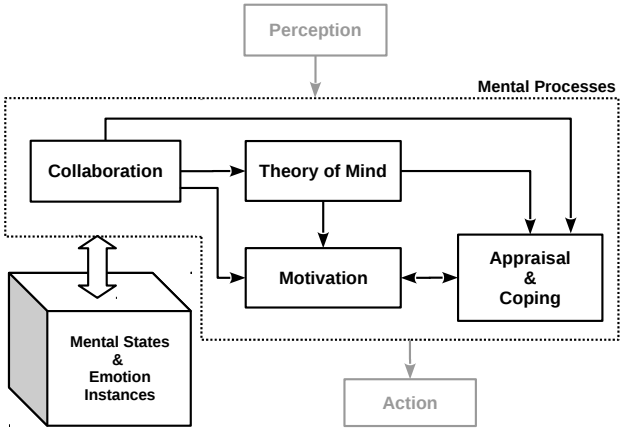


Figure 1: Computational framework based on Affective Motivational Collaboration Theory (arrows indicate primary influences between mechanisms).

Mental States

The Mental States shown in Figure 1 comprise the knowledge base required for all the mechanisms in the overall model.

Mental states are conscious states of the mind providing the content for cognitive processes. Affective Motivational Collaboration Theory operates with the following Mental States: beliefs, intentions, motives, goals and emotion instances. These Mental States possess attributes, each of which provides a discriminating and unique interpretation of the related cognitive entities. The self uses Mental States' attributes whenever there is an arbitration in the internal cognitive processes. The Appraisal mechanism and the Motivation mechanism play an essential role in computing the value of these attributes. I provide more details about these attributes in this section.

Belief: *Beliefs* are a crucial part of the Mental States. I have two different perspectives on categorization of beliefs. In one perspective, I categorize beliefs based on whether they are shared or not between the collaborators. The SharedPlans (Grosz and Sidner 1990) theory is the foundation of this categorization in which for any given proposition the agent may have: a) private beliefs (the agent believes the human does not know these), b) the inferred beliefs of the human (the agent believes the human collaborator has these beliefs), and c) mutual beliefs (the agent believes both the self and the human have these same beliefs and both of them believe that). From another perspective, I categorize beliefs based on who or what they are about. In this categorization, beliefs can be about the self, the other, or they can be about the environment. Beliefs about the environment can be about internal events, such as outcomes of a new appraisal or a new motivate, or external events such as the human's offer, question or request, and general beliefs about the environment in which the agent is situated. Beliefs can be created and updated by different processes. They also affect how these processes function as time passes.

The attributes of a belief are involved in arbitration procedures within different processes in the *Affective Motivational Collaboration Theory*. They impact a range of these processes from, the formation of new beliefs, the evaluation of an external event by the Appraisal mechanism, generation of new motives, updates on collaboration plan, to the activation of coping strategies and ultimately the self's behavior. The following six attributes of beliefs are most related to *Affective Motivational Collaboration Theory*.

- ***Strength:*** Belief strength is about how strong the self holds salient beliefs about an object, an entity, or an anticipated behavior. It can be measured through scales, for instance, how probable or likely that belief is, or just whether it is true or false. The strength of a belief can impact the self's intention attributes such as the *certainty* or *ambivalence*. A belief can be strong, but not accurate.
- ***Accuracy:*** Accuracy of a belief is the relation between that belief and the truth which that belief is about. The accuracy of a belief can be measured by looking at how closely that belief can relate to the truth. The accuracy of a belief as a gradational property can be used in evaluative processes of the self, i.e., Appraisal. It can also impact the self's other goal-driven processes, e.g., Motivation, by updating the utility function(s) with respect to the estimated belief accuracy.
- ***Frequency:*** The frequency of a belief is related to how regularly it appears as the result of an internal or an external event. The frequency of beliefs can impact attributes of the self's other Mental States. For instance, beliefs forming or maintaining intentions with direct experiences (see Section **??**) are more likely to occur frequently.
- ***Recency:*** The recency of a belief refers to how temporally close a particular belief is to the current state of collaboration. The recency attribute of the self's belief can bias (recency effect) the evaluation processes of the cognitive mechanism during collaboration. It can create a tendency to weight recent events more than earlier ones whenever it is required according to self's Mental States. The self can allow or hinder this tendency to adopt an appropriate Coping mechanism.
- ***Saliency:*** The saliency of a belief is a cognitive attribute that pertains to how easily the self becomes aware of a belief. This property of a belief has a prominent influence on the attention mechanism during collaboration. It directs the self's focus of attention to the most pertinent spatio-temporal salient internal or external event(s). Although belief saliency can determine the self's focus of attention, the self does not necessarily select an action based on the salient events.
- ***Persistence:*** It is argued that beliefs form and change due to cognitive and social considerations (Carley 1990). Persistent beliefs are very resistant to these changes. However, even persistent beliefs can change. Persistence of goal-related belief(s) influences the self's intentions and subsequently behaviors.

Motive: *Motives* are mental constructs which can initiate, direct and maintain goal-directed behaviors. They are created by the emotion-regulated Motivation mechanism. Motives can cause the formation of a new intention for the agent according to: a) its own emotional states (how the agent feels about something), b) its own private goal (how an action helps the agent to make progress), c) the collaboration goal (how an action helps to achieve the shared goal), and d) other's anticipated beliefs (how an action helps the other). Motives also possess a set of attributes, e.g., *Insistence* or *Failure Disruptiveness* (see Section **??**). These attributes are involved in comparison of newly generated motives based on the current state of the collaboration. Ultimately, the agent forms or updates a belief about the winning motive in the Mental States.

According to Sloman, motives can be compared on various dimensions (Sloman 1987). This comparison is based on motive attributes. In *Affective Motivational Collaboration Theory* motives are formed based on the self's existing Mental States under the influence of Appraisal mechanism. The existence of different Mental States, and the results of self appraisal as well as the reverse appraisal of the other

can cause a variety of motives to be formed. The Motivation mechanism needs a set of attributes to compare newly generated motives and choose the one which is most related to the current state of the collaboration. I have chosen the following five motive attributes as most related to the collaboration context.

- ***Insistence:*** The insistence of a motive defines the "interrupt priority level" of the motive, and how much that motive can attract the self's focus of attention. This dimension of motive is associated with what the Appraisal mechanism considers as *relevance* and *desirability* when evaluating an event. Beliefs about successive subgoals and the other's anticipated Mental States influence the insistence attribute of a motive. Insistent motives have higher priority and are able to interrupt self's ongoing tasks. The insistence of a motive is a function of the importance, urgency and the elapsed time for that motive.
- ***Importance:*** The importance of a motive is determined by the corresponding beliefs about the effects of achieving or not achieving the associated goal. It is a function of belief attributes (including strength, accuracy, frequency, recency, saliency and persistence) and the current goal. For instance, if a motive is supported by a belief about the current goal with relatively high attribute values, that motive will become important for the self.
- ***Urgency:*** The urgency of a motive defines how much time the self has to acknowledge and address that motive before it is too late. The urgency of a motive is a function of beliefs about the other's mental states as well as the required and the estimated time to fulfill the associated goal. For instance, the self responds to an urgent motive due to the existence of an important anticipated outcome for the other, and limited time to accomplish the corresponding tasks, even if those tasks are not important for the self.
- ***Intensity:*** The intensity of a motive determines how actively and vigorously that motive can help the self to pursue the goal if adopted. Motives with higher intensity will motivate the self to apply certain types of coping processes for an obstructed goal to avoid termination of the collaboration. Motives with higher intensity cause the self to find alternative solutions for the problem rather than abandoning the goal and ultimately the collaboration.
- ***Failure Disruptiveness:*** The failure disruptiveness attribute of a motive determines how disruptive failure is to achieving the corresponding goal. In other words, it gives the self a measure of the pleasantness of achieving a related goal. This attribute directs the self's behavior toward positive and negative outcomes during collaboration. It also plays a role in performance assessment processes when the self needs to compare its competence level on a given task relative to the other.

Intention: *Intentions* are mental constructs directed at future actions. They play an essential role in: a) taking actions according to the collaboration plan, b) coordination of actions with human collaborator, c) formation of beliefs about self and anticipated beliefs about the other, and d) behavior selection in the Coping mechanism. First, taking actions means that the agent will intend to take an action for primitive tasks that have gained the focus of attention, possess active motives, have satisfied preconditions for which required temporal predecessors have been successfully achieved. Second, intentions are involved in action coordinations in which the human's behavior guides the agent to infer an anticipated behavior of the human. Third, intentions play a role in belief formation mainly as a result of the permanence and commitment inherent to intentions in subsequent processes, e.g., appraisal of the human's reaction to the current action and self regulation. And lastly, intentions are involved in selecting intention-related strategies, e.g., planning, seeking instrumental support and procrastination, which these strategies are an essential category of the strategies in the Coping mechanism. Intentions possess a set of attributes, e.g. *Involvement, Certainty, Ambivalence* (see Section **??**) which moderate the consistency between intention and behavior. The issue of consistency between the intentions (in collaboration) and the behaviors (as a result of the Coping mechanism in the appraisal cycle) is important because neither of these two mechanisms alone provides solution for this concern.

The attributes of an intention influence several processes in *Affective Motivational Collaboration Theory*. They are involved in mechanisms such as Appraisal and Motivation as well as other Mental States, e.g., goals. One of the most important uses of intention attributes is to moderate the intention-behavior relations (Cooke and Sheeran 2004). Ultimately, the self can show more consistent behavior with respect to its own preceding behaviors and current state of the collaboration. I decided to include the following five intention attributes extracted from the psychology literature in *Affective Motivational Collaboration Theory*.

- ***Temporal Status:*** The temporal status of an intention can be defined as the extent to which an intention remains consistent over time. The self needs to maintain the stability of its intentions as time passes until the task is performed. Temporally stable intentions helps the other to accurately predict the self's behavior. The anticipated cognitive load of perceiving the self's task by the other impacts the temporal stability of the self's intention. In other words, the temporal stability of an intention moderates the intention-behavior relation of the self during collaboration.
- ***Direct Experience:*** The direct experience of an intention refers to whether the self previously has performed a task based on the similar intention. The self can refer to the corresponding Mental States of the intention directly experienced in the past before taking a new action. The Mental States associated with the prior experience of an intention can influence the appraisal of a new event requiring the self to perform the same task. For instance, the existence of a direct experience of an intention impacts the degree of the *expectedness* and *controllability* of an event during the collaboration which ultimately guides the Coping mechanism to produce an appropriate behavior.
- ***Certainty:*** The certainty of an intention is determined by

the quality of the underlying motive and the beliefs associated with that motive. The more *strong, accurate, frequent, recent, salient* and *persistent* a set of pertinent beliefs of the self are, the more chance the related motive has to be selected. Since the certainty of an intention depends on the associated motive, the nature of the pursued goal also implicitly impacts the certainty of that intention. A goal with higher *specificity* (see Section **??**) value influences the Belief Formation process of a new motive by increasing the certainty of the affiliated intention. The certainty of an intention is an important moderator of the self's intention-behavior consistency.

- ***Ambivalence:*** The Mental States of the self might contain contradictory intentions towards the pursuit of the same goal, which makes those intentions ambivalent. For instance, the self might already have an intention to perform a task according to the shared plan, while the Appraisal and the Motivation mechanisms dynamically provide a new motive forming a new opposing intention. Furthermore, ambivalent intentions can occur because of the contrast between the self's private goal and the shared goal during the collaboration. The ambivalence attribute of an intention is against the intention-behavior consistency of the self.
- ***Affective-Deliberative Consistency:*** The self's intentions possess an affective and a deliberative component. The affective component refers to the emotion instance and in general the affective evaluation of the self's intention towards its own behavior. However, the deliberate component refers to the self's actual intention which is formed either based on the existing shared plan or through a new motive generated by the Motivation mechanism. For instance, as an example of affective-deliberative inconsistency, the self can appraise the formation of the current intention as an *urgent* and *uncontrollable* one (which leads the self's emotion towards anger), despite the fact that performing the task associated with this intention is required for the satisfaction of the shared plan. In general, mutually consistent affective and deliberate components of an intention positively impacts the consistency of the self's intention and behavior.

Goal: *Goals* help the agent to create and update its collaboration plan according to the current private and shared goal content and structure, i.e., the *Specificity, Proximity* and *Difficulty* of the goal. Goals direct the formation of intentions to take appropriate corresponding actions during collaboration. Goals also drive the Motivation mechanism to generate required motive(s) in uncertain or ambiguous situations, e.g., to minimize the risk of impasse or to reprioritize goals. The *Specificity* of goals has two functions for the agent. First, it defines the performance standard for evaluating the progress and quality of the collaboration. Second, it serves the agent to infer the winner of competing motives. The *Proximity* of goals distinguishes goals according to how "far" they are from the ongoing task. Proximal (or short-term) goals are achievable more quickly, and result in higher motivation and better self-regulation than more temporally distant (or long-term) goals. Goals can influence the *Strength* of beliefs, which is an important attribute for regulating the elicitation of social emotions. The *Difficulty* of goals impacts collaborative events and decisions in the appraisal, reverse appraisal, motive generation and intention formation processes. For instance, overly easy goals do not motivate; neither are people motivated to attempt what they believe are impossible goals.

The attributes of a goal also impact the processes in *Affective Motivational Collaboration Theory*, especially the processes involved in Motivation and Appraisal mechanisms. The attributes of a goal are important because the Motivation and the Appraisal mechanisms in this theory are goal-driven and attribution of the goals according to the self's standards provides coherency of the processes and their outcomes. I discuss the three most relevant goal attributes in this section.

- ***Proximity:*** Goals can be distinguished by how far they project into the future during the collaboration. Proximal (short-term) goals result in more related motives and subsequently better self and social-regulation than temporally distant goals. Proximal goals impact the self's behaviors by influencing the evaluation process of the Appraisal mechanism. As a result, the self can determine the collaboration progress towards the shared goal more accurately while operating based on proximal goals.
- ***Specificity:*** Goals incorporating specific performance standards are more likely to enhance the self's self-evaluation than general goals. Specific goals raise the self-evaluation performance because they provide a more accurate baseline for the mechanisms, e.g., Appraisal or Motivation, that the self needs for self-evaluation during collaboration. Consequently, by increasing the self-evaluation performance, the self can compute more accurate anticipated self-satisfaction. For instance, holding an object A in a particular position with respect to an object B for a certain amount of time and welding them with a material C is a more specific goal than a general goal of installing an object A on object B.
- ***Difficulty:*** The goals that are moderately difficult have the most impact on the Motivation mechanism, and ultimately the self and social regulation processes of the self. Conversely, overly easy or impossible goals usually do not motivate an individual to achieve the goal. Difficult goals increase the probability of a motive's *failure disruptiveness*, and overly easy goals decrease the *importance* of the related motive; in both cases the motives have less chances to form new intentions. The lower chance of new intention formation will disrupt the self and social regulation since the self cannot regulate internal processes and influence the external world without forming appropriate intentions to take required actions. The existence of a partial shared plan, dependency on the other to perform a task, the failure of the same or similar task in the past and the conflict between the self's private goal and the shared goal all increase the difficulty level of a goal.

Emotion Instance: *Emotions* in Mental States are emotion instances that are elicited by the Appraisal mechanism (see Section **??** for list of emotion types used in this model).

The agent also keeps beliefs about these emotion instances in the Mental States. The Belief Formation mechanism creates or updates these beliefs about emotions. These emotion instances include the agent's own emotions as well as the anticipated emotions of the other which are created with the help of the processes in the Theory of Mind mechanism.

Each emotion has its own functionality in either intrapersonal or interpersonal level. These emotions not only regulate the self's internal processes, but also assist the self to anticipate the other's Mental States. In this section, I provide the description of some of the emotions that can be elicited during collaboration, and are involved in our scenario (see Section **??**). In this document, to avoid controversial issue of whether virtual agents or robots can feel emotions, I am going to use the convention of having emotions by the agent or the robot. The agent can also possess belief about an emotion instance which is similar to having belief about any other proposition.

- ***Joy:*** Joy is the state of an individual's well-being and is associated with the sense of successful achievement of a goal. Joy reveals one's sense of pleasure which implies an impending gain for the individual.
- ***Anger:*** Anger can be elicited by an unfair obstacle, hindering the individual's goal attainment and it is usually triggered by some external event (e.g., threat) which provokes a behavioral reaction. Anger functions to set boundaries or escape from dangerous situations, and implies an urgent desire for justice.
- ***Hope:*** Hope is the result of an optimistic evaluation of an event by an individual having expectations of positive and desirable future outcomes related to that event. It is usually a poignant assimilation of the present discontent and the future content implying an imagined or anticipated successful future goal state.
- ***Guilt:*** Guilt is based on self-condemnation in response to a negative outcome of one's self performance evaluation. It is caused by the violation of others' beliefs about the self, and others' standards and bearing significant responsibility due to that violation. The occurrence of guilt usually implies the desire to atone in social context.
- ***Pride:*** Pride is a product of the satisfied sense of one's own actions or decision outcomes. It implies the self-approval of the evaluation oucomes of one's own actions. Pride is associated with the achievement motivation (see Section **??**) wherein succeeding at a particular goal motivates the corresponding action.
- ***Shame:*** Shame is produced when one evaluates one's own actions or behaviors and attributes failure to the outcome. The individual focuses on specific features of self which led to failure. Shame implies the existence of remorse.
- ***Worry:*** Worry is one's emotional attempt to avoid anticipated potential threats or unidentified undesirable events. The individual's concern can be about a real or an imagined issue. Worry implies a fear of a future failure about which one should make a decision or take an action at present.

Collaboration Mechanism

Short paragraph to describe collaboration mechanism and its role in overall system (mention that we are not talking about the algorithms in this mechanism).

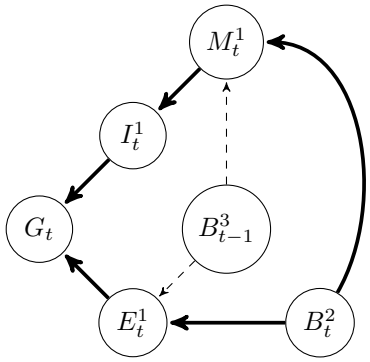
Then, describe the collaboration model and its components.

Appraisal Mechanism

Short paragraph to describe appraisal mechanism and its role in overall system.

Mental Graph

A graph illustration of mental state and a clear short and desciptive walkthrough example.



Appraisal Processes

A short paragraph to describe what variable we have chosen and why.

Relevance

Description + Algorithm + Example

Desirability

Description + Algorithm + Example

Expectedness

Description + Algorithm + Example

Controllability

Description + Algorithm + Example

Conclusion

Discussion about the application of the appraisal process in different mechanisms in our computational theory including Motivation mechanism.

References

Carley, K. 1990. On the persistence of beliefs. working paper. *Department of Social and Decision Sciences, Carnegie Mellon University*.

Cooke, R., and Sheeran, P. 2004. Moderation of cognition-intention and cognition-behaviour relations: A meta-analysis of properties of variables from the theory of planned behaviour. *British Journal of Social Psychology* 43:159–186.

Gratch, J., and C.Marsella, S. 2004. A domain-independent framework for modeling emotion. *Cognitive Systems Research* 5(4):269–306.

Grosz, B. J., and Sidner, C. L. 1990. Plans for discourse. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. Cambridge, MA: MIT Press. 417–444.

Sloman, A. 1987. Motives, mechanisms, and emotions. *Cognition & Emotion* 1(3):217–233.