Appraisal in Human-Robot Collaboration

Mahni Shayganfar, Charles Rich, Candace L. Sidner
Worcester Polytechnic Institute
Computer 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 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.

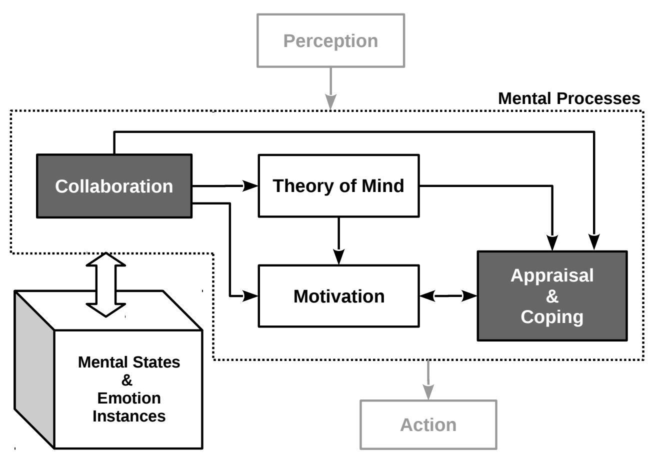


Figure 1: Computational framework based on Affective Motivational Collaboration Theory (arrows indicate primary influences between mechanisms).

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 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*. 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* 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* 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 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. 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 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.

The Collaboration mechanism is responsible for maintaining the internal structure of a collaboration, including the focus of attention, constraints on actions, updating the shared plan and, in general, monitoring of the collaboration. All of these structures require updating each time the self receives an external event. For instance, an utterance by the other can impact the self's focus of attention during the collaboration, or the effect of a primitive action can influence the self's view of an impasse on a task. As another example, the perception of the other's emotion instance can cause significant changes in the self's collaboration monitoring. The Collaboration mechanism also responds to internal events. These changes can occur under the influence of several other processes which are able to alter the self's mental states. For instance, the Motivation mechanism can generate a new motive which causes the self to add a new intention to its own Mental States.

- **Input:** The input to the *Collaboration* mechanism includes all the data that affects the execution of individual tasks in the collaboration plan. This data will be provided via the different elements of Mental States including beliefs, intentions and goals. These Mental States will establish agent's initial plan and will be updated throughout the collaboration process by the Perception and Inference mechanisms.
- **Output:** The output of *Collaboration* includes all the data that is modified or created during execution of a plan in the form of Mental States. These Mental States will be used by the internal processes in the Theory of Mind mechanism. Additionally, the Appraisal mechanism will use these Mental States to evaluate the events during collaboration. These Mental States also will be used by the Inference mechanism for the purpose of maintaining the collaboration.
- **Function:** The *Collaboration* mechanism will construct a hierarchy of tasks and also manage and maintain the constraints and other required details of the collaboration specified by the plan. These details include the inputs and outputs of individual tasks, the *preconditions* specifying whether it is appropriate to perform a task, and the *postconditions* specifying whether a just-completed task was successful (which can be used as an indication of an impasse or failure). *Collaboration* also keeps track of the focus of attention, which determines the salient objects, properties and relations at each point of the collaboration. Moreover, *Collaboration* has the ability to shift the focus of attention during the collaboration. All the other mechanisms in the overall *Affective Motivational Collaboration Model* are influenced by changes in the collaboration plan.

Appraisal Mechanism

Appraisal of External Events

The other's *utterances*, the effect(s) of the collaborators' *primitive actions*, and the other's *emotion instances* (expressed nonverbally) are the three types of external events perceived by the self during collaboration. The Appraisal mechanism receives the output of the Perception and Collaboration mechanisms as well as the requisite Mental States related to current event. It appraises the external event, in terms of appraisal variables using, a) the collaboration structure, b) the social characteristics of collaboration, and c) the history of the self's related Mental States. The collaboration structure contains information about the collaboration's shared plan and the collaborators' shared goal, the temporal and the hierarchical constraints of the tasks, and the current focus of attention.

The social characteristics of collaboration (see Section are conceptual properties associated with collaboration. These concepts provide deeper understanding of the collaborative interactions in terms of the underlying processes from both the personal and the social perspectives. Furthermore, these concepts determine when each process is used during a collaboration and why. Lastly, the self will progressively generate and update various types of Mental States during collaboration. The occurrence of a new external event causes a change in the self's Mental States. The construct of the new mental state, e.g., beliefs, are semantically connected to the older ones. The Appraisal mechanism uses the history of the Mental States to consistently evaluate a new external event.

Appraisal of Internal Events

The formation of a *new belief* (one that has not already been evaluated by the Appraisal mechanism, such as a belief about a new motive), and the formation of a *new motive*, are the two types of internal events that will be evaluated by the Appraisal mechanism. Thus, the Appraisal mechanism is involved in the formation of a new belief to ultimately create balanced behaviors during the course of a collaboration. The Appraisal mechanism receives the output of the Theory of Mind and Motivation mechanisms as well as the available Mental States (e.g., beliefs about the current emotion instances of the self and the other) and evaluates the environment in terms of the appraisal variables whenever a new belief is generated. As mentioned in Section, a) the collaboration structure, b) the social characteristic of the collaboration, and c) the history of the self's Mental States, are also crucial in the evaluation of internal events. The outcome of this appraisal process can result in a new belief about emotion instances, e.g. believing to have hope, or it can generate beliefs about different attributes of the Mental States of the self or the other.

Appraisal

- **Input:** The most significant part of *Appraisal's* input data is based on the activity of the Collaboration mechanism. This data includes all the required Mental States associated with the Collaboration mechanism. For instance, the beliefs and their Strengths will be used by algorithms inside of *Appraisal* to compute the value of the appraisal variables. *Appraisal* also receives data from the Theory of Mind mechanism. This data helps the agent use *Appraisal* for inferring the human's intentions and motives based on a reverse appraisal procedure.

  The input data from the Perception is generally needed to support the evaluation of the events. The Motivation mechanism sends data to *Appraisal* whenever, for example, a new set of motives is being generated or the motive comparator wants to evaluate different motives for an arbitration procedure.

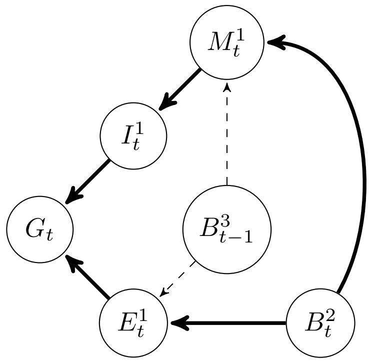
- **Output:** The output of *Appraisal* can directly and indirectly impact other mechanisms. The Motivation mechanism uses this data to generate, compare and monitor motives based on the current internal appraisal of the agent as well as the appraisal of the environment. This output data also assists the Belief Formation mechanism to create new beliefs about the current emotional state of the agent or the human as well as the corresponding appraisal of individual motives.

- **Function:** *Appraisal* is a subjective evaluation mechanism based on individual processes each of which computes the value of the appraisal variables used in my computational model. The Collaboration mechanism needs the evaluative assistance of the *Appraisal* mechanism for various reasons. The course of a collaboration is based on a full or a partial plan which needs to be updated as time passes and collaborators achieve, fail at or abandon a task assigned to them. The failure of a task should not destroy the entire collaboration. Appraising the environment and the current events helps the agent to update the collaboration plan and avoid further critical failures during collaboration. *Appraisal* also helps the agent to have a better understanding of the human's behavior by making inferences based on appraisal variables. Furthermore, in order to collaborate successfully, a collaborator cannot simply use the plan and reach to the shared goal; there should be an adaptation mechanism not only for updating the plan but also the underlying Mental States. For instance, there are beliefs about the appraisal of the self and the other that augment the model of what collaborators have done, and what and how they are planning to achieve the current shared goal based on their emotional states. This process will be discussed in more detail under the Motivation mechanism. Additionally, the beliefs formed based on the appraisals can impact other mechanisms such as the Theory of Mind, Motivation and Coping, essentially including the whole computational model.

Mental Graph

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

The *FINDPATH* function uses the Dijkstra algorithm to find the shortest path.



Appraisal Processes

All of the algorithms use a mental graph which is a directed acyclic graph constructed based on the mental states of the robot at each turn during the collaboration (see Section). They also use the content of the most recent event at each turn which includes the associated goal and the belief about the event. The associated goals can be any type action, utterance, or emotional expression.

We found four appraisal variables, i.e., *Relevance* (Algorithm 1), *Desirability* (Algorithm 2), *Expectedness* (Algorithm 3), and *Controllability* (Algorithm 4) to be more important than the other appraisal variables in a collaboration context. There are other appraisal variables introduced in psychological (Scherer, Schorr, and Johnstone 2001) and computational literature (Gratch and Marsella 2004); we do not provide algorithms for them in this paper. The concept behind these variables are either too simple, e.g., *Perspective* (i.e., the view point from which an event is judged), or it requires a particular approach which is not our concern in this paper, e.g., *Likelihood* (i.e., how probable is the outcome of an event).

The *recognizeGoal()* function returns the unique goal to which the event directly contributes irrespective of the event type (i.e., action, utterance, or emotional expression). This event structure holds in all of our algorithms.

Relevance

Relevance as an appraisal variable measures the significance of an event for the robot. An event can be evaluated to be relevant if it has a positive utility or it can causally impact a state with a positive utility (Marsella and Gratch 2009). Relevance is an important appraisal variable since the other appraisal dimensions are only derived for the relevant events.

Algorithm 1 determines the relevance of the given event with respect to the shared goal. An event is *irrelevant* if there is no connection between the event and the current shared goal. If there is a connection, the relevance of the event will depend on the significance of the event with respect to the current collaboration status. The significance of an event is determined based on the utility of the event as it is also presented in (Gratch and Marsella 2004; Marsella and Gratch 2009). We believe although the utility of the event represents the sigificance of the event, the other collaborator's expressed emotion also plays a role by influencing the significance of the utility through a threshold. As a result, evaluating the relevance of the events can cause a collaborative robot to effectively respond only to the events which can positively impact the status of the shared goal without dedicating all other resources to every single event. The relevance process also benefits the information that the collaboration structure contains, e.g., shared goal.

The algorithm starts by taking the belief about the current event (b_{ε_t}) and the shared goal (g_t) associated with the current mental graph. After perceiving an event, it is the belief about that event which represents the event in robot's mental state. Continuing in line 3, the g_t represents the shared goal (in the mental graph) at time (turn) t within the shared plan. Then, we find the shortest path ($\mathcal{P}$) between the corresponding nodes of b_{ε_t} and g_t in the mental graph. As we mentioned earlier, if there is no path between the belief about the current event and the active shared goal, the algorithm finds the event *irrelevant* to the current shared goal, and terminates.

However, if there is a path between b_{ε_t} and g_t, we need a way to determine whether the event is relevant to the current collaboration status. Therefore, first, we compute the utility of the event ($\mathcal{U}$) such that ($0 \leq \mathcal{U} \leq 1$) based on the values of the attributes associated with the belief about that event (b_{ε_t}), as well as the attributes of the motive within the path $\mathcal{P}$. Yet, the significance of an event in a collaborative environment not only is based on the utlity of the event, but it is also influenced by the emotional state of the other collaborator. In other words, the relevance of a belief about an event is influenced by the perceived emotion of the human collaborator. The human's emotion influences the decision about the utility of the event in form of a threshold value (τ_t). For instance, a positive expressed emotion of the human reduces the threshold value which consequently makes the robot find an event relevant with even a slightly positive utility.

We compute this threshold value (τ_t) using a Fuzzy Logic system on a three-dimensional space of the somatic markers associated to the other's expressed emotion as they are described in (Breazeal 2002). In this space every emotion can be mapped to a vector of three values, i.e., *Arousal, Valence* and *Stance*. The details about this process is out of this paper's context. Finally, we make our decision about the relevance of an event with respect to the human's emotional state. Consequently, an event can be considered *irrelevant* even though there is a path between b_{ε_t} and g_t.

Desirability

Desirability characterizes the value of an event to the robot in terms of whether the event facilitates or thwarts the collaboration goal. Desirability captures the valence of an event with respect to the robot's preferences (Gratch and Marsella 2004). In a collaborative robot preferences are biased towards those events facilitating progress in the collaboration. An event is desirable if it somehow facilitates the state of the shared goal, or if it inhibits the state of a goal that is inconsistent with respect to the shared goal.

Desirability plays an important role in the overall architecture; it makes the processes involved in the other mechanisms (e.g., Motivation or Theory of Mind) and consequently the robot's mental state congruent with the collabo-

Algorithm 1 (Relevance)

function IsEventRelevant(MentalGraph $\mathcal{G}_t$, Event ε_t)
  $b_{\varepsilon_t} \leftarrow$ getBelief(ε_t)
  $g_t \leftarrow$ *currGoal*$(\mathcal{G}_t)$
  $\mathcal{P} \leftarrow$ findPath(b_{ε_t}, g_t)
  if $(\mathcal{P} = \emptyset)$ **then**
    return IRRELEVANT
  else
    $\mathcal{U} \leftarrow$ getEventUtility(b_{ε_t}, g_t)
    $\tau_t \leftarrow$ getEmotionalThreshold$(\mathcal{G}_t)$
    if $(\mathcal{U} \geq \tau_t)$ **then**
      return RELEVANT
    else
      return IRRELEVANT

ration status which is the robot's desire. Therefore, it causes the robot to dismiss events causing inconsistencies in the robot's collaborative behavior. Moreover, desirability is also crucial from the collaboration's point of view. A collaborative robot needs to know whether its own and the other collaborator's actions, utterances, and emotional expressions are desirable in terms of their consistence with the status of the current shared goal. In other words, the collaboration mechanism uses the appraisal process of desirability to coordinate what the self or the other does, says, and expresses during collaboration. Reciprocally, the appraisal mechanism and in this case the desirability process use the collaboration structure to obtain their required information.

Algorithm 2 provides a process in which the desirability of an event is computed with regard to the status of the shared goal; i.e., it operates based on whether and how the event changes the status of the current shared goal. It receives the current mental graph, $\mathcal{G}_t$, and the current event, ε_t, from input, and decides whether and how the event is desirable or undesirable. First, the algorithm checks the status of the collaboration's top level goal (lines 2 to 6), and if the the top level goal is still in progress, it continues by checking the status of the current shared goal (lines 7 to 13). If any of the top level and current shared goals are achieved in these two steps, the robot interprets the event as a desirable one. However, if any of these goals are blocked, the event will be considered undesirable by the robot.

The algorithm continues in the case of an unknown status of the current shared goal, and checks whether the precondition(s) of the associated goal with the current event, g_{ε_t}, are satisfied (lines 17 to 21). The robot prefers the satisfied preconditions and interprets the event as desirable while unsatisfied preconditions are undesirable for the robot. For instance, a satisfied precondition of a future goal is still desirable for the robot to some extent. Note that the robot also checks the ambiguity of the associated goal with the current event (line 15). An ambiguous goal is a goal which is not recognized in the robot's plan, and it is undesirable for the robot. After all, if the preconditions of the associated goal with the current event are unknown, the robot checks whether this goal, g_{ε_t}, contributes to the current shared goal, g_t (lines 22 to 25). As a result a contributing goal will obtain a neutral desirability in comparison with a noncontributing goal which will be undesirable for the robot.

Algorithm 2 (Desirability)

1: **function** IsEventDesirable(MentalGraph $\mathcal{G}_t$, Event ε_t)
2:   **if** (*topLevelGoalStatus()* = ACHIEVED) **then**
3:     **return** HIGHEST-DESIRABLE
4:   **else if** (*topLevelGoalStatus()* = BLOCKED) **then**
5:     **return** HIGHEST-UNDESIRABLE
6:   **else if** (*topLevelGoalStatus()* = INPROGRESS) **then**
7:     **if** (*currGoalStatus()* = ACHIEVED) **then**
8:       **return** HIGH-DESIRABLE
9:     **else if** (*currGoalStatus()* = BLOCKED) **then**
10:       **return** HIGH-UNDESIRABLE
11:     **else if** (*currGoalStatus()* = INPROGRESS) **then**
12:       **return** MEDIUM-DESIRABLE
13:     **else if** (*currGoalStatus()* = UNKNOWN) **then**
14:       $g_{\varepsilon_t} \leftarrow$ *recognizeGoal*(ε_t)
15:       **if** $(g_{\varepsilon_t} =$ AMBIGUOUS$)$ **then**
16:         **return** HIGH-UNDESIRABLE
17:       **if** (*precondStatus(*g_{ε_t}*)* = SATISFIED) **then**
18:         **return** LOW-DESIRABLE
19:       **else if** (*precondStatus(*g_{ε_t}*)* = UNSATISFIED) **then**
20:         **return** LOW-UNDESIRABLE
21:       **else if** (*precondStatus(*g_{ε_t}*)* = UNKNOWN) **then**
22:         $g_t \leftarrow$ *currGoal*$(\mathcal{G}_t)$
23:         **if** $($*doesContribute*$(g_{\varepsilon_t}, g_t))$ **then**
24:           **return** NEUTRAL
25:         **else if** $(\neg$*doesContribute*(g_{ε_t}, g_t) **then**
26:           **return** MEDIUM-UNDESIRABLE

Expectedness

Expectedness is the extent to which the truth value of a state could have been predicted from causal interpretation of an event (Marsella and Gratch 2009). In the collaboration context the expectedness of an event measures the congruency of the event with respect to the existing knowledge about the shared goal. Therefore, expected events are those of which beliefs about them are congruent to the status of the collaboration since their associated goals are expected with respect to the shared plan.

Expectedness underlies a collaborative robot's attention by evaluating the congruence of events with respect to the structure of an existing shared plan. Congruent beliefs in a robot's mental state will lead to more consistent and effective outcomes of the processes in the overall architecture. Therefore, a collaborative robot uses expectedness to maintain its own mental state towards the shared goal. The robot will also be able to respond to unexpected but relevant events. As a result, the collaboration mechanism uses expectedness to maintain the robot's attention and subsequently its mental state with respect to the shared goal. In parallel, the appraisal mechanism uses the underlying information of the collaboration structure to evaluate the expectedness of an event.

Algorithm 3 (Expectedness)

1: **function** IsEventExpected(MentalGraph $\mathcal{G}_t$, Event ε_t)
2:   $g_t \leftarrow$ *currGoal*$(\mathcal{G}_t)$
3:   $g_{\varepsilon_t} \leftarrow$ *recognizeGoal*(ε_t)
4:   **if** $(g_{\varepsilon_t} = \text{AMBIGUOUS})$ **then**
5:    **return** UNEXPECTED
6:   **if** $(g_t = g_{\varepsilon_t})$ **then**
7:    **return** EXPECTED
8:   **else**
9:    **if** $(\neg$*isAchieved*$(g_t))$ **then**
10:     **return** UNEXPECTED
11:    **else**
12:     **if** $(\neg$*isAchieved*$(g_t.parent))$ **then**
13:      **if** $($*doesContribute*$(g_{\varepsilon_t}, g_t.parent))$ **then**
14:       **return** EXPECTED
15:      **else**
16:       **return** UNEXPECTED
17:     **else**
18:      **if** $($*doesContribute*$(g_{\varepsilon_t}, g_t))$ **then**
19:       **return** EXPECTED
20:      **else**
21:       **if** $($*isFocused*$(g_{\varepsilon_t}))$ **then**
22:        **return** EXPECTED
23:       **else**
24:        **return** UNEXPECTED

In algorithm 3 we provide the process of the expectedness based on the shared plan and status of the shared goal. The key point in this algorithm is the status of the current shared goal (g_t) and its relationship with the goal associated with the current event (g_{ε_t}). The algorithm receives the current mental graph, $\mathcal{G}_t$, and the current event, ε_t, from input, and decides whether the current event is expected.

First, we need to extract the goal in the current mental graph and the recognized goal associated with the current event. Similar to the desirability algorithm (Algorithm 2), we check whether the g_{ε_t} is ambiguous. In the case of ambiguity in g_{ε_t}, we consider the current event unexpected since an effective collaboration requires perceivable and unambiguous goals asscoiated with the events. We continue by the comparison of the current shared goal and the recognized goal associated with the current event with respect to the shared plan. If these two goals are not the same, it is possible that the current shared goal is already achieved. The event will be unexpected (line 10) if the current shared goal is not achieved and the current event does not refer to the same goal. However, if the current goal is achieved, it is important to see whether its parent is also achieved (line 12). This step is important because the event can be expected if the new goal contributes to the parent of the recently achieved goal. Therefore, if the parent goal in the hierarchical plan is not achieved, the contribution of the associated goal to the current event can help us to decide whether the event is expected (lines 12 to 16). However, if the parent goal is already achieved, the new goal can contribute (as a child) to the recently achieved shared goal, i.e., g_t, which is also expected (line 19). On the contrary, if the new goal does not contribute to g_t, it might be a goal in another branch in the shared plan which has received focus and should be achieved. In such a case, again, the event will be expected; otherwise we consider the event unexpected (lines 21 to 24).

Controllability

Controllability is the extent to which an event can be influenced, and it is associated with a robot's ability to cope with the appraised event (Gratch and Marsella 2004). Thus, the robot can determine whether the outcome of an event can be altered by some actions under either of the collaborators' control. In other words, controllability is a measure of a robot's ability to maintain or change a particular state as a consequence of an event.

Controllability is also important for the overall architecture. For instance, the robot can choose to ask or negotiate about a collaborative task which is not controllable; it can cause the robot to interpret or predict the other's emotional state (e.g., anger if the task is blocked, i.e., uncontrollable for the other), or form a new motive to establish an alternative goal for the current uncontrollable event. In general, other mechanisms in the architecture use the appraisal process of controllability in their decision making processes; meanwhile the controllability uses the information from the collaboration structure, e.g., successful predecessors of a goal.

Algorithm 4 (Controllability)

function IsEventControllable(Event ε_t)
  $\alpha \leftarrow$ GetAgencyRatio(ε_t)
  $\beta \leftarrow$ GetAutonomyRatio(ε_t)
  $\lambda \leftarrow$ GetSucPredecessorsRatio(ε_t)
  $\mu \leftarrow$ GetAvailableInput(ε_t)
  $\mathcal{U} \leftarrow \frac{\omega_0 \cdot \alpha + \omega_1 \cdot \beta + \omega_2 \cdot \lambda + \omega_3 \cdot \mu}{\omega_0 + \omega_1 + \omega_2 + \omega_3}$
  $\tau_t \leftarrow$ getEmotionalThreshold()
  if $(\mathcal{U} \geq \tau_t)$ **then**
   return CONTROLLABLE
  else
   return UNCONTROLLABLE

An important determinant of one's emotional response is the sense of control over the occurring events. This sense of subjective control is based on one's reasoning about self's power. For instance, the robustness of one's plan for executing actions can increase sense of power and subsequently the sense of control. In the collaboration context, we have translated the sense of control into a combination of four different factors including a) *agency* and b) *autonomy* of the robot, as well as the ratios of c) *successful predecessors*, and d) the *available inputs* of a given goal (i.e., g_{ε_t}) in the shared plan.

In Algorithm 4, we compute the controllability of an event based on these four factors (lines 2 to 5). Algorithms 5 to 8 are used to illustrate the undelying processes of all these factors. We use weighted averaging over these four factors to compute the utility of an event in terms of controllability of the event. However, the value of all these weights are set to 1.0 for the purpose of simplicity at this stage of the project. We will use these weights after investigating more about the influence of these factors, and implementing other mechanisms in the overall architecture. After computing the value of the utility, we compare this value to an emotional threshold similar to what we discussed in Algorithm 1. This comparison leads to our decision about the controllability of an event (lines 8 to 11).

Algorithm 5 (Get Agency Ratio)

```
function GETAGENCYRATIO(Event ε_t)

    b_{ε_t} ← GETBELIEF(ε_t)
    g_{ε_t} ← recognizeGoal(ε_t)

    if (g_{ε_t} = AMBIGUOUS) then
        return 0.0

    P ← FINDPATH(b_{ε_t}, g_{ε_t})

    if (P ≠ ∅) then

        M_{g_t} ← currMotive(P)

        if (M_{g_t} ≠ ∅) then
            if (M_{g_t} · type = INTERNAL) then
                return 1.0
            else
                return 0.0
        else
            return 0.0
    else
        return 0.0
```

Agency is the capacity of an individual to act independently in any given environment. In a collaborative environment sometimes collaborators require to act independently of each other. Hence, they need to have some internal motives that are formed based on their own mental state rather than being reinforced by the other. These internal motives will lead the collaborators to acquire new intentions towards new goals whenever it is required. In Algorithm 5, if there is a path between the belief about the current event and the recognized goal associated with that event, we exploit the motive within the acquired path. We consider maximum agency value denoted as α in Algorithm 4 (i.e., $\alpha = 1.0$) if the robot's mental state possesses an internal motive towards the recognized goal; otherwise we consider the minimum agency value (i.e., $\alpha = 0.0$) for no or external motives. Note that the process of forming new internal motives is out of this paper's context.

Autonomy is the ability of making decisions without the influence of the others. Autonomy implies acting on one's own and being responsible for that. In a collaborative environment, tasks are delegated to the collaborators based on their capabilities. Therefore, each collaborator is responsible for the delegated task and the corresponding goal. In Algorithm 4, β denotes the value of autonomy with regard to the event (ε_t). This value is the ratio of the number of the contributing goals to g_{ε_t} for which the robot is responsible over the total number of contributing goals to g_{ε_t} (see Algorithm 6). If the associated goal with the current event corresponds to a nonprimitive task, the algorithm checks the responsible agent for each primitive task contributing to the nonprimitive one and returns a value of which ($0 \leq \beta \leq 1$). However, if the associated goal to the current event corresponds to a primitive task the value of β would be 0 or 1. In general, higher autonomy leads to more positive value of the controllability.

Algorithm 6 (Get Autonomy Ratio)

```
function GETAUTONOMYRATIO(Event ε_t)

    g_{ε_t} ← recognizeGoal(ε_t)

    if (g_{ε_t} = AMBIGUOUS) then
        return 0.0

    Φ_g ← extractContributingGoals(g_{ε_t})

    for all φ_g^i ∈ Φ_g do
        if (getResponsible(φ_g^i) = SELF) then
            count_self ← count_self + 1

    return count_self / Φ_g.total()
```

The structure of a shared plan accommodates the order of the required predecessors of a goal. Predecessors of a goal, g, are other goals that the collaborators should achieve before trying to achieve goal g. In Algorithm 7, we use the ratio of successfully achieved predecessors of the recognized goal (g_{ε_t}) associated with the current event over the total number of predecessors of the same goal. This ratio (denoted as λ in Algorithm 4) is another factor to compute controllability of an event. If all of the predecessors of the given goal are already achieved, then $\lambda = 1$ which is the maximum value for λ. On the contrary, failure of all of the predecessors will lead to $\lambda = 0$. Therefore, higher λ value positively impacts the value of controllability for the current event.

Inputs of a task are the required elements that the collaborators use to achieve the specified goal of the task. These inputs are also part of the structure of a shared plan. In Algorithm 8, we extract the required inputs of the associated goal

Algorithm 7 (Get Succeeded Predecessors Ratio)

function GETSUCPREDECESSORSRATIO(Event ε_t)

  $g_{\varepsilon_t} \leftarrow$ *recognizeGoal*(ε_t)

  if $(g_{\varepsilon_t} =$ AMBIGUOUS$)$ **then**
    return 0.0

  $\Theta g \leftarrow$ *extractPredecessors*(g_{ε_t})

  for all $\theta_g^i \in \Theta_g$ **do**
    if $($*isAchieved*$(\theta_g^i))$ **then**
      $count_{achieved} \leftarrow count_{achieved} + 1$

  return $count_{achieved}$ / $\Theta g.total()$

with the current event, and check whether all the required inputs are available for the goal g_{ε_t}. The outcome will be the ratio of the available required inputs over the total required inputs of the goal associated with the current event. This value (denoted as μ in Algorithm 4) will be bound to 0 and 1. Similar to the other factors in controllability process, the closer value of μ gets to 1, it has more positive impact on overall controllability value of the event.

Algorithm 8 (Get Available Input Ratio)

function GETAVAILABLEINPUTRATIO(Event ε_t)

  $g_{\varepsilon_t} \leftarrow$ *recognizeGoal*(ε_t)

  if $(g_{\varepsilon_t} =$ AMBIGUOUS$)$ **then**
    return 0.0

  $\mathcal{X}_g \leftarrow$ *extractInputs*(g_{ε_t})

  for all $\chi_g^i \in \mathcal{X}_g$ **do**
    if $($*IsAvailable*$(\chi_g^i))$ **then**
      $count_{available} \leftarrow count_{available} + 1$

  return $count_{available}$ / $\mathcal{X}_g.total()$

Conclusion & Future Work

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

References

Breazeal, C. 2002. *Designing Sociable Robots*. MIT Press.

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 Marsella, S. C. 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.

Marsella, S. C., and Gratch, J. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10(1):70–90.

Scherer, K. R.; Schorr, A.; and Johnstone, T. 2001. *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press.

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