Appraisal in Human-Robot Collaboration

Mahni Shayganfar, Charles Rich, Candace L. Sidner
Worcester Polytechnic Institute, Worcester Massachusetts
mshayganfar | rich | sidner @wpi.edu

Abstract

We have investigated the mutual influence of affective and collaborative processes in a cognitive theory to support interaction between humans and robots or virtual agents. We have developed new algorithms for appraisal processes, as part of a new overall computational model for implementing collaborative robots and agents. We build primarily on the *cognitive appraisal* theory of emotions and the *SharedPlans* theory of collaboration to investigate the structure, fundamental processes and functions of emotions in a collaboration. We have evaluated our implemented algorithms by conducting an online user study.

1 Introduction

Sousa in The Rationality of Emotion [Sousa, 1990] makes the case for claiming that humans are capable of rationality largely because they are creatures with emotions. The idea of having robots or other intelligent agents living in a human environment has been a persistent dream from science fiction books to artificial intelligence and robotics laboratories. Collaborative robots are expected to become an integral part of humans' environment to accomplish their industrial and household tasks. In these environments humans will be involved in robots' operations and decision-making processes. The involvement of humans influences the efficiency of robots' interaction and performance, and makes them dependent on humans' cognitive abilities and mental states.

This work is implemented as part of a larger effort to build robots capable of generating and recognizing emotions in order to be better collaborators. In this paper, we report on the specific problem of appraising events within a collaborative interaction. Our contribution is to ground general appraisal concepts in the specific context and structure of collaboration. This work is part of the development of *Affective Motivational Collaboration Theory* which 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].

After discussing related works, we briefly introduce the Affective Motivational Collaboration Theory, focusing on the collaboration and appraisal mechanisms as well as mental states. We then provide more details about the graph representation of the robot's mental state. Next, we describe the algorithms we developed to compute the value of four crucial appraisal variables. To compare the results from our algorithms with humans' decisions we have conducted a user study using crowd sourcing; the results are provided in Section 7.

2 Related Work

Our work builds on the general notions of appraisal theory [Gratch and Marsella, 2004; Marsella *et al.*, 2010; Scherer, 1999; Scherer *et al.*, 2001], but is focused on its application in human-robot collaboration. Computational appraisal models have been applied to a variety of uses including psychology, robotics, AI, and cognitive science. For instance, in [Marsella and Gratch, 2009] EMA is used to generate specific predictions about how human subjects will appraise and cope with emotional situations. Furthermore, appraisal theory has also been used in robots' decision making [Gonzalez *et al.*, 2013], or in their cognitive systems [Hudlicka, 2007; Marinier III and Laird, 2008]. Additionally, in the virtual agents community, empathy and affective decision-making is a research topic that has received much attention in the last two decades [McQuiggan and Lester, 2007; Paiva *et al.*, 2004; Pontier and Hoorn, 2013; Velàsquez, 1997]. However, EMA and several other examples in artificial intelligence and robotics which apply appraisal theory do not focus on the dynamics of collaborative contexts [Adam and Lorini, 2014; Kim and Kwon, 2010; Marsella and Gratch, 2009; Rosenbloom *et al.*, 2015].

The computational collaboration model in our work is strongly influenced by the SharedPlans theory [Grosz and Sidner, 1990]. However, our algorithms are also compatible with other collaboration theories, e.g., Joint Intentions theory [Cohen and Levesque, 1991], or STEAM [Tambe, 1997]. These theories have been extensively used to examine and describe teamwork and collaboration. Yet, collaboration and emotion theories have never been combined, as they are in our work. We believe a systematic integration of collaboration theories and appraisal theory can help us describe the underlying collaboration processes leading to the existing collaboration structures.

3 Affective Motivational Collaboration Theory

Affective Motivational Collaboration Theory deals with the interpretation and prediction of observable behaviors in a dyadic collaboration [Shayganfar *et al.*, 2016]. 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 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. The theory explains how emotions regulate the underlying processes when these events occur. It also elucidates the role of *motives* as goal-driven emotion-regulated constructs with which a robot can form new intentions to cope with events.

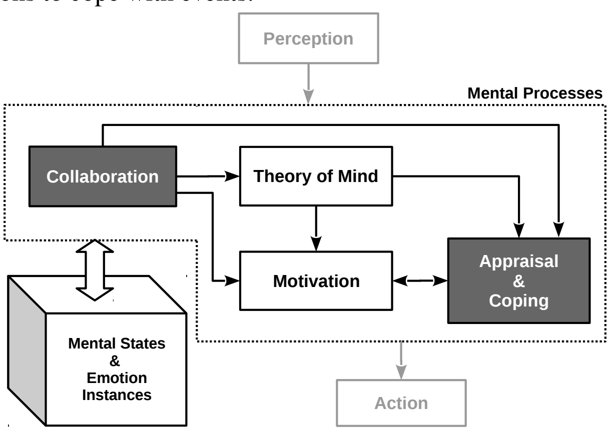

Figure 1: Computational framework based on Affective Motivational Collaboration Theory (arrows indicate primary influences between mechanisms).

Our focus is on the mechanisms depicted as mental processes in Figure 1 along with the mental states. Each mechanism includes one or more processes in our architecture. For instance, the *Collaboration* mechanism includes processes such as *Focus Shifting* and *Constraint Management*, while as we discuss in Section 6 the *Appraisal* mechanism includes processes to compute the values for different appraisal variables. 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 (see Figure 2). 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 infers a model of the other's anticipated mental state. The self progressively updates this model during the collaboration.

3.1 Mental States

A brief description of mental states is provided as prerequisite knowledge for understanding the appraisal processes. 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 mind providing the content for cognitive processes. These mental states possess attributes, each of which provides a unique interpretation of the related cognitive entities. The self uses these attributes whenever there is an arbitration in the internal cognitive processes. We only describe some of the attributes of beliefs and motives in this paper, since they are used in our appraisal algorithms.

Beliefs are a crucial part of the mental states. Beliefs have attributes and they impact different processes of the framework such as the evaluation of an external event by the Appraisal mechanism, and updates to the collaboration plan. We use three belief attributes in in Appraisal mechanism. Belief *strength* is about how strongly the self holds salient beliefs about an object, an entity, or an anticipated behavior. The *saliency* of a belief is a cognitive attribute that pertains to how easily the self becomes aware of a belief. The *persistence* of a belief refers to how resistant the belief is to changes.

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 robot according to: a) its own emotional states, b) its own private goal, c) the collaboration (shared) goal, and d) other's anticipated beliefs. Motives possess a set of attributes. The Motivation mechanism compares motives based on the quality of these attributes and chooses the one which is the most related to the current state of the collaboration. We use two motive attributes in Appraisal mechanisms. The *importance* of a motive is determined by the corresponding beliefs about the effects of achieving or not achieving the associated goal. The *urgency* of a motive defines how much time the self has to acknowledge and address that motive before it is too late.

Intentions are mental constructs directed at goals and future actions. They play an essential role in taking actions according to the collaboration plan as well as behavior selection in the Coping mechanism. Intentions are also involved in selecting intention-related strategies, e.g., planning, seeking instrumental support and procrastination.

Goals help the robot to create and update its collaboration plan according to the current private and shared goal content and structure. Goals direct the formation of intentions to take appropriate corresponding actions during collaboration.

Emotions in mental states are emotion instances that are elicited by the Appraisal mechanism, e.g., *Joy, Anger, Hope, Worry*. These emotion instances include the robot'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.

4 Example Scenario

The example scenario is part of a much larger interaction we are implementing to test our theory. This example shows a very short part of an interaction between a robot and an astronaut during their collaboration. Their mission is to finish installing a few solar panels together. However, the astronaut encounters a measurement tool problem:

Astronaut [turn t-1]: Oh no! Finishing the quality check of our installation with this measurement problem is so frustrating. I think we should stop now!

Robot [turn t]: I see. This is frustrating. But, I can help you with the measurement tool and we can finish the task as originally planned.

In this scenario, the robot appraises the problem with the measurement tool as a *relevant, undesirable, unexpected*, but *controllable* event. Consequently, the coping mechanism first acknowledges the astronaut's negative valenced emotion (i.e., frustration), then provides a new plan to continue the collaboration.

5 Collaboration

The Collaboration mechanism constructs a hierarchy of goals associated with tasks in the form of a hierarchical task network (see Figure 2), and also manages and maintains the constraints and other required details of the collaboration including 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). Collaboration also keeps track of the focus of attention, which determines the salient objects, properties and relations at each point, and shifts the focus of attention during the interaction.

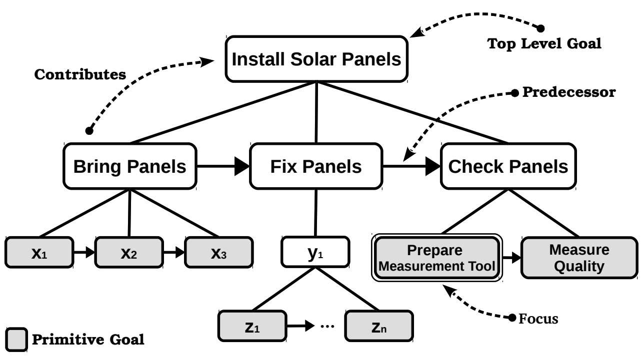

Figure 2: Collaboration structure (shared plan).

Here, we briefly describe the methods which retrieve information about the collaboration structure, and are used in our algorithms to compute the values of appraisal variables. In these methods, ε_t is the event corresponding to time *t*, and g_t is a given goal at time *t*.

- *recognizeGoal(ε_t)* returns the unique goal to which the given event (action, utterance, or emotional expression) directly contributes, or *ambiguous* if this method does not recognize a goal in the plan.
- *topLevelGoalStatus(g_t)* returns the status of the top level goal whether it is ACHIEVED, FAILED, BLOCKED, INAPPLICABLE, PENDING, or IN PROGRESS. In our example, "Install Solar Panels" is the top level goal.
- *currGoalStatus(g_t)* returns the current goal status whether it is ACHIEVED, FAILED, BLOCKED, INAPPLICABLE, PENDING, or IN PROGRESS. In our example, "Prepare Measurement Tool" is the current (focused) goal.
- *precondStatus(g_t)* returns the status of the precondition for the given goal whether it is SATISFIED, UNSATISFIED or UNKNOWN. For instance, the precondition for fixing a panel is whether the panel is appropriately located on its frame.
- *isLive(g_t)* returns *true* if all the predecessors of the given goal are ACHIEVED and all the preconditions of the goal are SATISFIED; otherwise returns *false*.
- *isFocusShift(g_t)* returns *true* if the given goal is not the previous focus (top of the stack); otherwise returns *false*.
- *isNecessaryFocusShift(g_t)* returns *true* if the status of the previous focus was ACHIEVED; otherwise returns *false* [Lesh *et al.*, 2001].
- *isPath(g_1, g_2)* returns *true* if there is a path between g_1 and g_2 in a plan tree structure; otherwise returns *false*.
- *doesContribute(g_t)* returns whether the given goal contributes to another goal in the higher level of the plan hierarchy. For instance, an abstract (nonprimitive) goal of "Bring Panels" contributes to the higher level goal of "Install Solar Panels".
- *extractContributingGoals(g_t)* returns all the contributing goals of the given goal. For instance, "Prepare Measurement Tool" and "Measure Quality" are two goals contributing to the "Check Panels" nonprimitive goal.
- *extractPredecessors(g_t)* returns the predecessors of the given goal. For instance, the "Fix Panels" goal is the predecessor of another goal called "Check Panels".
- *extractInputs(g_t)* returns all the required inputs for the given goal. For example, the goal "Fix Panels" requires inputs such as the *welding tool* and the *panel*.
- *isAvailable(g_t)* returns whether the given input is available. For instance, if the *welding tool* is required for the goal "Fix Panels", is it available now?
- *isAchieved(g_t)* returns whether the given goal is achieved, i.e., whether all the postconditions of the given goal are SATISFIED.
- *isFocused(g_t)* returns whether the focus is on given goal now. In this example, the focus is on the goal "Prepare Measurement Tool". The focused goal is the goal that the robot is currently pursuing.
- *getResponsible(g_t)* returns responsible agents of the given goal. In a dyadic collaboration, both of the agents can be partly responsible for a nonprimitive goal, while each is responsible for one or more primitive goals. For instance, both the robot and the astronaut are responsible for the nonprimitive goal of "Install Solar Panels", whereas it is only the astronaut who is responsible for the primitive goal of "Prepare Measurement Tool".

6 Appraisal Processes

We consider four appraisal variables to be the most important appraisal variables in a collaboration context, i.e., *Relevance* (Algorithm 1), *Desirability* (Algorithm 2), *Expectedness* (Algorithm 3), and *Controllability* (Algorithm 4). There are other appraisal variables introduced in psychological [Scherer *et al.*, 2001] and computational literature [Gratch and Marsella, 2004]. We believe most of these variables can be straightforwardly added to our appraisal mechanism later. All of the algorithms in this section use mental states of the robot (discussed in Section 3.1) which are formed based on the collaboration structure. These algorithms use the corresponding recognized goal of the most recent event at each turn.

6.1 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 variables are more meaningful only for relevant events.

Algorithm 1 determines the relevance of the given event with respect to the current mental state. The relevance of the event depends 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 significance of the event, the other collaborator's expressed emotion also plays a role by influencing the significance of the utility through a threshold value. As a result, evaluating the relevance of the events can cause a collaborative robot to respond effectively to the events which can positively impact the status of the shared goal, without dedicating all resources to every single event. The relevance process also benefits from the information that the collaboration structure contains, e.g., shared goal.

Algorithm 1 (Relevance)

function IsEventRelevant(Event ε_t)

  $g_t \leftarrow$ *recognizeGoal*(ε_t)

  $\mathcal{U} \leftarrow$ getEventUtility(g_t)
  $\tau_t \leftarrow$ getEmotionalThreshold(g_t)

  if $(\mathcal{U} \geq \tau_t)$ **then**
    return RELEVANT
  else
    return IRRELEVANT

After perceiving an event, it is the belief about that event which represents the event in the robot's mental state. Also, *recognizeGoal* returns the goal (g_t) to which the current event contributes, unless it is *ambiguous*; g_t represents the shared goal at time (turn) t within the shared plan. We compute the utility ($0 \leq \mathcal{U} \leq 1$) of the event based on the values of the attributes associated with the existing beliefs in the mental state, as well as the attributes of the motive associated with the recognized goal. We use three of the belief attributes discussed in Section **??** to compute the belief related part of the utility:

- *Strength*: The extent to which the pre and postconditions of a goal and its predecessors and/or contributing goals are SATISFIED or UNSATISFIED makes a belief about the goal stronger. Respectively, an UNKNOWN pre and postcondition status of a goal and its predecessors and/or contributing goals forms beliefs with lower strength.
- *Saliency*: Beliefs related to the goal at the top of the focus stack are more salient than beliefs related to any other goal in the plan, whether those goals are already ACHIEVED or FAILED, or they will be pursued in the future.
- *Persistence*: The recurrence of a belief over the passage of time (turns) increases the persistence of the belief. Beliefs occurring only in one turn have the lowest value of persistence.

We also use two of the motive attributes discussed in Section **??** to compute the motive related part of the utility ($\mathcal{U}$):

- *Urgency*: There are two factors impacting the urgency of a motive: a) whether the goal directing the given motive is the predecessor of another goal for which the other collaborator is responsible, and b) whether achieving the goal directing the given motive can mitigate the other collaborator's negative valenced emotion.
- *Importance*: A motive is important if failure of the directing goal causes an impasse in the shared plan (i.e., no further goal is available to achieve), or achievement of the directing goal removes an existing impasse.

We compute the utility of an event based on these five attributes. The value of each attribute is between 0 and 1, and we consider the same weight for each attribute. These weights can be learned or modified when our framework is fully implemented. The value of the overall utility is computed using a simple weighted averaging function which results in an overall value between 0 and 1.

The significance of an event in a collaborative environment is based not only on the utility of the event, but it is also influenced by the perceived emotion of the human collaborator. The human's emotion influences the decision about the utility of the event in the form of a threshold value τ_t (see Algorithm 1). For instance, a positively expressed emotion of the human reduces the threshold value which consequently makes the robot find an event relevant with even a slightly positive utility. This threshold value (τ_t) is currently determined based on whether the valence of the human's perceived emotion is positive (e.g., happiness) or negative (e.g., anger). Consequently, an event can be considered IRRELEVANT even though the utility has a relatively positive value, because relevance is influenced by the human's perceived emotional state.

6.2 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. Desirability plays an important role in the overall architecture; it makes the processes involved in the other mechanisms (e.g., Motivation and Theory of Mind), and consequently the robot's mental state, congruent with the collaboration status which is a collaborative 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.

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 distinguishes between the top level goal and the current goal because the top level goal's change of status attains a higher positive or negative value of desirability. For instance, failure of the top level goal (e.g., installing solar panel) is more undesirable than failure of a primitive goal (e.g., measuring the quality of the installed panel).

An AMBIGUOUS goal is a goal associated with the current event (ε_t) which is not recognized in the robot's plan; therefore it is UNDESIRABLE for a collaborative robot. A top level goal' status must be ACHIEVED (i.e., SATISFIED postcondition) to consider the event MOST-DESIRABLE. When the goal's status is FAILED (i.e., UNSATISFIED postcondition) or BLOCKED, the associated event has the MOST-UNDESIRABLE or UNDESIRABLE values respectively. A goal is BLOCKED if any of the required goals or goals recursively through the parent goal are not ACHIEVED. An INAPPLICABLE goal is also considered as UNDESIRABLE. A goal is INAPPLICABLE if any of its predecessors are not ACHIEVED, and/or its preconditions are not SATISFIED. For PENDING and INPROGRESS top level goals, the status of the current goal associated with the top level goal determines the status of the event ε_t. Only a non-primitive goal can have INPROGRESS status, if it has been started but is not yet completed. A goal can be PENDING if it is live, or if it is a non-primitive goal that has not been started yet. ACHIEVED current goals mark an event (ε_t) as DESIRABLE, while FAILED or BLOCKED current goals render the event associated with them as MOST-UNDESIRABLE and UNDESIRABLE respectively. PENDING or INPROGRESS current goals mark their associated events as NEUTRAL.

Algorithm 2 (Desirability)

```
function IsEventDesirable(Event ε_t)

    g_t ← recognizeGoal(ε_t)

    if (g_t = AMBIGUOUS) then
        return UNDESIRABLE

    if (topLevelGoalStatus(g_t) = ACHIEVED) then
        return MOST-DESIRABLE
    else if (topLevelGoalStatus(g_t) = FAILED) then
        return MOST-UNDESIRABLE
    else if (topLevelGoalStatus(g_t) = BLOCKED)
    (topLevelGoalStatus(g_t) = INAPPLICABLE) then
        return UNDESIRABLE
    else if (topLevelGoalStatus(g_t) = PENDING)
    (topLevelGoalStatus(g_t) = INPROGRESS) then

        if (currGoalStatus(g_t) = ACHIEVED) then
            return DESIRABLE
        else if (currGoalStatus(g_t) = FAILED) then
            return MOST-UNDESIRABLE
        else if (currGoalStatus(g_t) = BLOCKED)
        (topLevelGoalStatus(g_t) = INAPPLICABLE) then
            return UNDESIRABLE
        else if (topLevelGoalStatus(g_t) = PENDING)
        (currGoalStatus(g_t) = INPROGRESS) then
            return NEUTRAL
```

6.3 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 evaluates the congruency of the event with respect to the existing knowledge about the shared goal. Thus, expectedness underlies a collaborative robot's attention. Congruent beliefs in a robot's mental state will lead to more consistent and effective outcomes of the processes in the overall architecture. The collaboration mechanism uses expectedness to maintain the robot's attention and subsequently its mental state with respect to the shared goal. Reciprocally, the appraisal mechanism uses the underlying information of the collaboration structure to evaluate the expectedness of an event. 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.

In Algorithm 3 we provide the process of computing 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) that is associated with the event ε_t and its relationship with the top level goal (g_{top}).

The intuition captured here is that one expects the current goal to be finished before undertaking another activity, but the goals that are the next focus of attention are also to be expected [Lesh *et al.*, 2001]. Therefore, if the goal is live, the algorithm checks whether the goal has not changed, or the interpretation of the last event results in a necessary focus shift. Shifting the focus to a new goal is necessary when the former goal is achieved and a new goal is required. Consequently the new event is the MOST-EXPECTED one. However, even if the focus shift is not necessary, the new event can be considered as EXPECTED, since the corresponding goal is already live. For goals that have not yet been started (that is, are not live), the algorithm must determine how unexpected it would be to pursue one now; if the goal is at least in the plan, i.e., on the path to the top level goal, it is just UNEXPECTED while any others are MOST-UNEXPECTED.

6.4 Controllability

Controllability is the extent to which an event can be influenced, and it is associated with a robot's ability to cope with

Algorithm 3 (Expectedness)

function IsEventExpected(Event ε_t)

$g_t \leftarrow$ *recognizeGoal*(ε_t)
$g_{top} \leftarrow$ *getTopLevelGoal*(g_t)

if (*isLive*(g_t)) **then**
  if ($\neg$*isFocusShift*(g_t)
   isNeccessaryFocusShift(g_t)) **then**
    return MOST-EXPECTED
  else
    return EXPECTED
else
  if (*isPath*(g_t, g_{top})) **then**
    return UNEXPECTED
  else
    return MOST-UNEXPECTED

an appraised event [Gratch and Marsella, 2004]. Thus, a 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.

Algorithm 4 (Controllability)

1: **function** IsEventControllable(Event ε_t)

2: $\alpha \leftarrow$ GetAgencyRatio(ε_t)
3: $\beta \leftarrow$ GetAutonomyRatio(ε_t)

4: $\lambda \leftarrow$ GetSucPredecessorsRatio(ε_t)
5: $\mu \leftarrow$ GetAvailableInput(ε_t)

6: $\mathcal{U} \leftarrow \frac{\omega_0 \cdot \alpha + \omega_1 \cdot \beta + \omega_2 \cdot \lambda + \omega_3 \cdot \mu}{\omega_0 + \omega_1 + \omega_2 + \omega_3}$

7: $\tau_t \leftarrow$ getEmotionalThreshold$()$

8: **if** $(\mathcal{U} \geq \tau_t)$ **then**
9:   **return** CONTROLLABLE
10: **else**
11:   **return** UNCONTROLLABLE

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, or the robot can 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 controllability uses the information from the collaboration structure, e.g., successful predecessors of a goal.

An important determinant of one's emotional response is the sense of control over the events occurring. 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 one's 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_t) in the shared plan.

In Algorithm 4, we compute the controllability of an event based on these four factors (lines 2 to 5). We use weighted averaging over these four factors to compute the utility of an event in terms of controllability of the event. The value of all these weights are set to 1.0 for the purpose of simplicity at this stage of the project. We will adjust these weights after further investigating 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 in Algorithm 4).

Agency is the capacity of an individual to act independently in any given environment. In a collaborative environment collaborators are sometimes required to act independently of each other. Hence, they need to have some internal motives that are formed based on their own mental states rather than motives that are reinforced by the other collaborator. These internal motives will lead the collaborators to acquire new intentions towards new goals whenever it is required. We extract the motive associated with the current goal in the mental state. We consider a 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 motives or external motives only. Note that the process of forming new internal motives is beyond scope of this paper.

Autonomy is the ability to make decisions without the influence of 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 goals contributing to g_t for which the robot is responsible over the total number of contributing goals to g_t. If the goal associated with the current event corresponds to a nonprimitive goal, the algorithm checks the responsible agent for each primitive goal contributing to the nonprimitive one and returns a value of which $(0 \leq \beta \leq 1)$. However, if the associated goal of the current event corresponds to a primitive goal the value of β would be 0 or 1. In general, higher autonomy leads to a more positive value of controllability.

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. We use the ratio of successfully achieved predecessors of the recognized goal (g_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 the third factor used to compute the 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, a higher λ value positively impacts the value of controllability for the current event.

Finally, ***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. We extract the required inputs of the associated goal with the current event, and check whether all the required inputs are available for the goal g_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 the controllability process, the closer the value of μ gets to 1, the more positive impact it has on the overall controllability value of the event.

In summary, the output of these four appraisal processes serves as critical input for the other mechanisms of the Affective Motivational Collaboration Framework, shown in Figure 1. By providing adequate interpretation of events in the collaborative environment, the appraisal mechanism enables the robot to carry out proper collaborative behaviors.

7 Evaluation

We developed our user study to test our hypothesis that humans will provide similar answers as our algorithms to questions related to different factors used to compute four appraisal variables. We conducted a between subject user study using an online crowdsourcing website – CrowdFlower[1]. We had one group of subjects for each questionnaire corresponding to an appraisal variable. There were 12 questions (including 2 test questions) in the controllability and expectedness questionnaires, 14 questions (including 2 test questions) in the desirability questionnaire, and 22 questions (including 3 test questions) in the relevance questionnaire. Each group originally had 40 subjects. To increase the quality of our subjects' answers, we limited the visibility of our questionnaires to a few English speaking countries, i.e., United States, Britain, and Australia. We also limited our subject pools to those that have acquired the highest confidence level on the crowdsourcing website. Our questionnaires included 2 or 3 test questions (depending on the length) to check the sanity of the answers. We eliminated subjects providing wrong answers to our sanity questions. We also eliminated subjects with an answering time less than 2 minutes. The final number of accepted subjects in each group is provided in Table 1.

Table 1: Evaluation Results

appraisal variables	# of subjects	mean	stdev	p-value
Relevance	29	0.713	0.107	¡0.001
Desirability	35	0.778	0.150	¡0.001
Expectedness	33	0.785	0.120	¡0.001
Controllability	33	0.743	0.158	¡0.001

[1] http://www.crowdflower.com

To minimize the background knowledge necessary for our test subjects, we used a simple domestic example of preparing a peanut butter and jelly sandwich, and a hard boiled egg sandwich for a hiking trip. We provided clear textual and graphical instructions for all four questionnaires. The instructions presented a sequence of hypothetical collaborative tasks to be carried out by the test subject and an imaginary friend, Mary, in order to accomplish their goal of preparing two sandwiches. Figure 3 shows the corresponding task model for these instructions. Test questions introduced specific situations related to the shared plan; these situations included, among others, blocked tasks, and failure or achievement of a shared goal provided in the instruction. Each question provided three possible answers (which were counterbalanced in the questionnaire). One option provided a distinct alternative; another option was used to provide a dichotomy with the first alternative, and a third option was used to check whether the subjects perceived the other two options as equal. We also provided a brief description as well as a simple example for each appraisal variable, e.g., *relevance*, at the end of the corresponding instructions. Using this approach, we prepared four different online questionnaires for the appraisal variables: *relevance, desirability, expectedness* and *controllability*. Note that the collaboration structure and the instructions were the same for all four questionnaires.

Each question was designed based on different factors that we use in our algorithms (see Section 6). Here, we present three example questions from the expectedness, controllability, and desirability questionnaires, and describe how each question relates to a specific factor within the corresponding algorithm. The input for our algorithms was the task model depicted in Figure 3.

Imagine you have pressed the two slices of bread (one covered with strawberry jam and one covered with peanut butter) together and passed it to Mary. Which of the following two actions is **more expected**?

A. Mary puts the given sandwich into a zip lock bag after cutting it in half.

B. Mary puts some pickles on another slice of bread.

C. Equally expected.

Figure 4: Example Expectedness Question.

Figure 4 shows the example question from the expectedness questionnaire. In this example, with respect to Algorithm 3 (line 6), option A is more expected because the task related to this option provides the next available task in the focus stack (see the task model in Figure 3). Although the task in option B is part of the existing task model, it is considered as unexpected by our algorithm, since it is not live in the plan. We provided option C to determine whether the human subjects will similarly differentiate between these two options. This question was presented to the human subjects to determine whether their decision for the expectedness of this event is similar to the output of the expectedness algorithm. For this question, the human decision was 97% similar to the algorithm's output. Average results for the expectedness questionnaire are presented in Table 1.

Figure 5 shows an example question from the controllability questionnaire. The algorithm's output is option B, and is

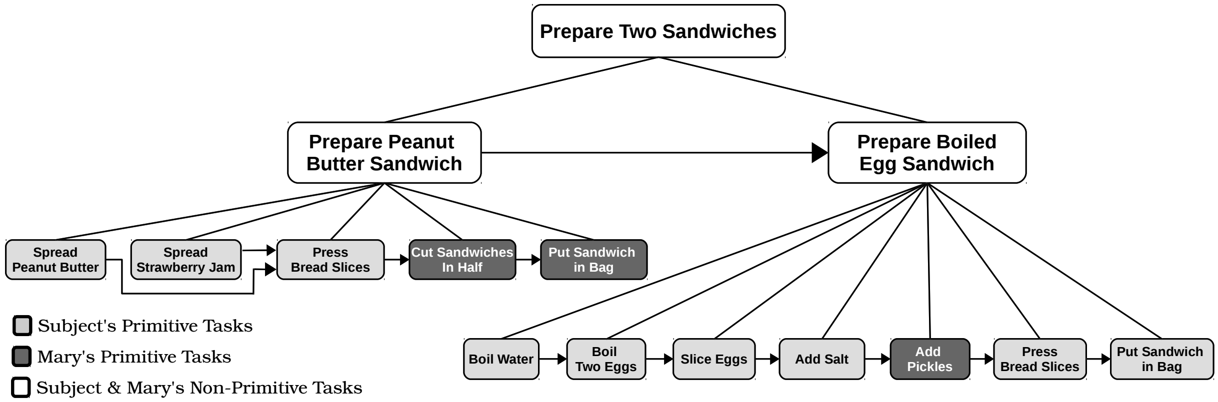

Figure 3: Collaboration Task Model for the Evaluation.

Imagine you want to make a peanut butter sandwich. Which of the following two actions is **more controllable**?

A. You can spread the peanut butter on one slice of bread and you need Mary to spread strawberry jam on the second slice of bread.

B. You can spread the peanut butter on one slice of bread and strawberry jam on the second slice of bread.

C. Equally controllable.

Figure 5: Example Controllability Question.

determined by Algorithm 4 (line 3), similarly to the expectedness example above. In this example, option B is more controllable than option A, because the self over total ratio of the responsibility of the predecessors of the given task (see *Autonomy* in Section 6.4) is higher than the ratio in option A; i.e., self is responsible to spread peanut butter on one slice of bread and strawberry jam on another slice of bread. In this question, the humans decision was 90% in agreement with the algorithm's output.

Which of the following two actions is **more desirable**?

A. Imagine you pressed two slices of bread together with peanut butter and strawberry jam on them, and passed them to Mary. Mary cuts the peanut butter sandwich in half and puts them in the zip lock bag.

B. Imagine you want to make the egg sandwich. You have sliced the eggs, put them on one slice of bread, salted them, and waiting for Mary to put some pickles on your eggs. Mary puts some pickles on your eggs.

C. Equally desirable.

Figure 6: Example Desirability Question.

Figure 6 shows an example question from the desirability questionnaire. The output based on the Algorithm 2 (line 14) is option C, since in both option A and option B, the focus goal has been achieved successfully. Therefore, in this example, both options A and B are desirable. The humans decision was 77% in agreement with the algorithm's output in this question.

We conducted the user study to compare the results with the implemented algorithms discussed in Section 6. As we mentioned, each question had 3 answers. Therefore, a totally random distribution would result in 33% agreement with our algorithms results. However, the average ratio indicating similarity between human subjects decisions and the output of our algorithms is significantly higher than 33%. The total number of subjects' answers similar to a) the *relevance* algorithm (n=29) averaged 71.3% (s=10.7%), b) the *desirability* algorithm (n=35) averaged 77.8% (s=15.0%), c) the *expectedness* algorithm (n=33) averaged 78.5% (s=12.0%), and d) the *controllability* algorithm (n=33) averaged 74.3% (s=15.8%). It is worth noting that the human subjects agreed 100% on some questions, while one some other questions there was a much lower level of agreement.

The results indicate that our algorithms provide appraisal variable outputs sufficiently similar to the decisions of human's appraisal. Our hypothesis in our evaluation was that our algorithms would correctly predict the judgements of humans on doing these tasks. Our results indicate that people largely performed as our hypothesis predicted. The p-values obtained based on a one-tailed z-test (see Table 1) show the probability of human subjects' data being generated from a random set. The very small p-values indicate that the data set is not random; in fact, the high percentage of similarity shows that the four appraisal algorithms predicted the human judgments.

8 Conclusion

While these results support our hypothesis, they are not a perfect prediction. We believe the difference between these two results could be due to several reasons, including: a) the fact that we conducted our study online and had little control on our subjects, b) our algorithms may require further granularity, or c) the difference between decision making processes of individuals, which can be affected by other factors such as personality, gender, and culture. While it may be possible to achieve a higher level of agreement between humans and the algorithms results, these results indicate that the current algorithms are adequate to be used in a collaboration context. In our future work, we will implement the remaining mechanisms in Affective Motivational Collaboration framework and carry out an end-to-end user study to verify the behavior of a collaborative robot using our architecture.

Acknowledgments

This work is supported by the National Science Foundation under award IIS-1012083. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[Adam and Lorini, 2014] Carole Adam and Emiliano Lorini. A BDI emotional reasoning engine for an artificial companion. In *Workshop on Agents and multi-agent Systems for AAL and e-HEALTH (PAAMS)*, volume 430, pages 66–78. Springer, 2014.

[Cohen and Levesque, 1991] Philip Cohen and Hector J. Levesque. *Teamwork*. SRI International, 1991.

[Gonzalez *et al.*, 2013] Alvaro Castro Gonzalez, Maria Malfaz, and Miguel Angel Salichs. An autonomous social robot in fear. *IEEE Transactions Autonomous Mental Development*, 5(2):135–151, 2013.

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

[Grosz and Sidner, 1990] Barbara J. Grosz and Candace L. Sidner. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 417–444. MIT Press, Cambridge, MA, 1990.

[Hudlicka, 2007] Eva Hudlicka. Reasons for emotinos: Modeling emotinos in integrated cognitive systems. In Wayne D. Gary, editor, *Integrated Models of Cognitive Systems*, volume 59, pages 1–37. New York: Oxford University Press, 2007.

[Kim and Kwon, 2010] Hyoung-Rock Kim and Dong-Soo Kwon. Computational model of emotion generation for human-robot interaction based on the cognitive appraisal theory. *Journal of Intelligent and Robotic Systems*, 60(2):263–283, 2010.

[Lesh *et al.*, 2001] Neal Lesh, Charles Rich, and Candace L. Sidner. Collaborating with focused and unfocused users under imperfect communication. In Mathias Bauer, Piotr J. Gmytrasiewicz, and Julita Vassileva, editors, *User Modeling 2001*, volume 2109, pages 64–73. Springer Berlin Heidelberg, 2001.

[Marinier III and Laird, 2008] Robert P. Marinier III and John E. Laird. Emotion-driven reinforcement learning. In *CogSci 2008*, 2008.

[Marsella and Gratch, 2009] Stacy C. Marsella and Jonathan Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, March 2009.

[Marsella *et al.*, 2010] Stacy C. Marsella, Jonathan Gratch, and Paola Petta. Computational models of emotion. In *Blueprint for Affective Computing (Series in Affective Science)*. Oxford University Press, Oxford, 2010.

[McQuiggan and Lester, 2007] Scott W. McQuiggan and James C. Lester. Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies*, 65(4):348–360, 2007.

[Paiva *et al.*, 2004] Ana Paiva, Joao Dias, Daniel Sobral, Ruth Aylett, Polly Sobreperez, Sarah Woods, Carsten Zoll, and Lynne Hall. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 194–201, 2004.

[Pontier and Hoorn, 2013] Matthijs Pontier and Johan F. Hoorn. How women think robots perceive them - as if robots were men. In *International Conference on Agents and Artificial Intelligence (ICAART-2)*, pages 496–504, 2013.

[Rosenbloom *et al.*, 2015] Paul S. Rosenbloom, Jonathan Gratch, and Volkan Ustun. Towards emotion in sigma: From appraisal to attention. In Jordi Bieger, Ben Goertzel, and Alexey Potapov, editors, *Artificial General Intelligence*, volume 9205 of *Lecture Notes in Computer Science*, pages 142–151. Springer International Publishing, 2015.

[Scherer *et al.*, 2001] Klaus R. Scherer, Angela Schorr, and Tom Johnstone. *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press, 2001.

[Scherer, 1999] Klaus R. Scherer. On the sequential nature of appraisal processes: Indirect evidence from a recognition task. *Cognition & Emotion*, 13(6):763–793, 1999.

[Shayganfar *et al.*, 2016] Mahni Shayganfar, Charles Rich, and Candace L. Sidner. An overview of appraisal in human-robot collaboration. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems – AAMAS (Extended Abstract)*, 2016.

[Sousa, 1990] Ronald De Sousa. *The Rationality of Emotion*. MIT Press, 1990.

[Tambe, 1997] Milind Tambe. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.

[Velàsquez, 1997] Juan D. Velàsquez. Modeling emotions and other motivations in synthetic agents. In *Proceedings of the 14th Nnational Conference on Artificial Intelligence AAAI-97*, pages 10–15, 1997.