Relevance and Controllability Appraisals in Human-Robot Collaboration

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

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 the relevance and controllability of 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].

There are several appraisal models contributing in different applications such as robotics, virtual agents and social sciences. However, none of these models have focused on the appraisal processes during collaboration. We believe appraisal plays a key role in collaboration due to its regulatory and evaluative nature. Also, collaboration induces some differences to underlying appraisal processes due to its unique nature. For instance, although the major appraisal models use utility to compute the relevance of an event, we have found new cognitive components involved in determining utility because of the influence of the collaboration. These components, such as recurrence of a belief by the human collaborator or influence of the human collaborator's peceived emotion on the robot's decisions emphasize the fact that collaboration creates demands for different procedures in appraisal processes. In this paper, we provide the relevance and controllability appraisals of an event within the collaboration context.

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 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 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.*, 2016c]. 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.

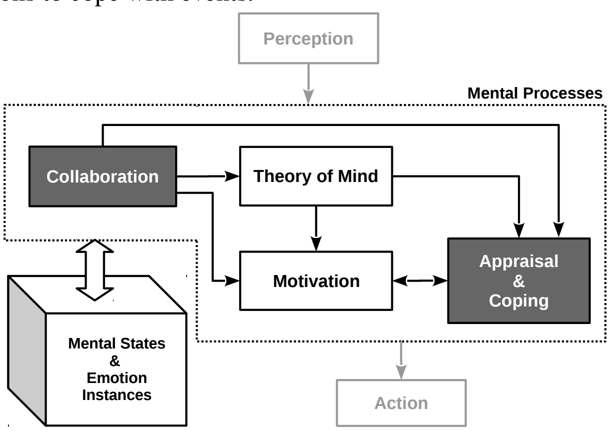

Figure 1: Computational framework based on Affective Motivational Collaboration Theory (arrows indicate primary influences between mechanisms).

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 5 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, and 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 the 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.

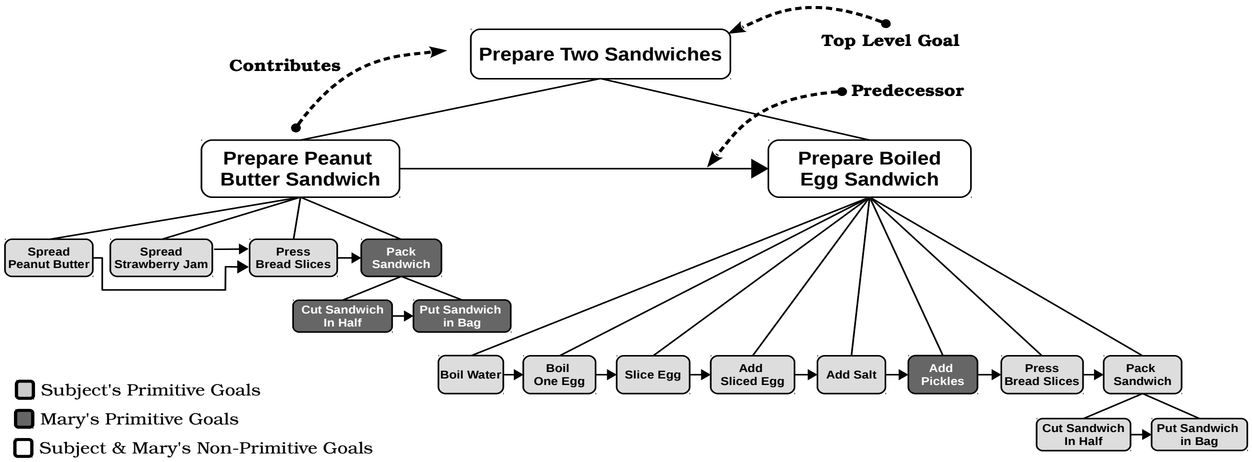

Figure 2: Collaboration structure (also used as task model for the evaluation).

4 Collaboration

The Collaboration mechanism constructs a hierarchy of goals associated with tasks in a hierarchical task network (see Figure 2), and also 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 monitors the focus of attention, which determines the salient objects, properties and relations at each point, and shifts the focus of attention during the interaction.

Here, we describe the methods which retrieve information about the collaboration structure, and are used 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; it is only one goal since the robot can only do one primitive action at a time in our collaboration model. However, in the goal tree, a given primitive action can contribute to more than one higher level (parent) goal. The method returns *ambiguous* if it does not recognize a goal in the plan.
- *currGoalStatus*(g_t) returns whether the current goal status is ACHIEVED, FAILED, BLOCKED, INAPPLICABLE, PENDING, or IN PROGRESS. In our example, "Add Pickles" 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 slicing the eggs is whether the eggs are boiled appropriately.
- *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*.
- *extractContributingGoals*(g_t) returns all contributing goals of the given goal. For instance, "Boil Water", "Slice Egg" and other goals in this level are multiple goals contributing to the "Prepare Boiled Egg Sandwich" nonprimitive goal.
- *extractPredecessors*(g_t) returns the predecessors of the given goal. For instance, the "Spread Peanut Butter" and "Spread Strawberry Jam" goals are the predecessors of another goal called "Press Bread Slices".
- *extractInputs*(g_t) returns all required inputs for the given goal. For example, the goal "Boil Water" requires inputs such as *Pot* and *Stove*.
- *isAvailable*(g_t) returns whether the given input is available. For instance, if the *Pot* is required for the goal "Boil Water", is it available now?
- *isFocused*(g_t) returns whether the focus is on the given goal. The focused goal is the goal that the robot currently pursues.
- *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 Mary and the subject are responsible for the nonprimitive goal "Prepare Peanut Butter Sandwich", whereas only Mary is responsible for the primitive goal "Spread Peanut Butter".

5 Appraisal Processes

We discuss two appraisal variables in a collaboration context, i.e., *Relevance* (since other appraisals are only derived for relevant events), and *Controllability* (since it is associated with the agent's coping ability). There are other appraisal variables introduced in psychological [Scherer *et al.*, 2001] and computational literature [Gratch and Marsella, 2004]. We have implemented other appraisal variables such as *expectedness* [Shayganfar *et al.*, 2016c] and *desirability* [Shayganfar *et al.*, 2016a] which do not appear in this paper due to space limitations. The algorithms in this section use mental states of the robot (see Section 3.1) which are formed based on the collaboration structure.

5.1 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 non-zero utility [Marsella and Gratch, 2009]. Relevance is an important appraisal variable since the other appraisal variables are more meaningful only for relevant events. However, the utility of an event during collaboration is influenced by the other collaborator's actions and mental states because there is a commitment between collaborators to achieve the shared goal based on the shared plan. Other appraisal models only consider the utlity of an event based on the self's (robot's) goal and plan.

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 collaboration status, which is determined based on the utility of the event as presented in [Gratch and Marsella, 2004; Marsella and Gratch, 2009]. We believe computing the relevance of an event during collaboration involves other factors which other appraisal models do not consider. For instance, the human's perceived emotion, recurrence of a belief, or occurrence of a belief about an unrelated goal by the human also play important roles by influencing the utility of an event during collaboration. As a result, evaluating the relevance of the events can cause a collaborative robot to respond effectively which can positively impact the status of the shared goal, without dedicating all resources to every event.

Algorithm 1 (Relevance)

```
function IsEventRelevant(Event ε_t)
    g_t ← recognizeGoal(ε_t)
    U ← getEventUtility(g_t)
    τ_t ← getEmotionalThreshold(g_t)
    if (τ_t ≤ |U|) then
        return RELEVANT
    else
        return IRRELEVANT
```

After perceiving an event, the belief about that event represents the event in the robot's mental state. *recognizeGoal* returns the goal (g_t) to which the current event contributes, unless it is *ambiguous*; g_t is the shared goal at time (turn) t within the shared plan. We compute the utility ($-1 \leq \mathcal{U} \leq 1$) of the event using the values of the attributes associated with the existing beliefs, and the attributes of the motive associated with the recognized goal. We use three belief attributes (see Section 3.1) to compute the belief-related part of the utility:

- *Strength (T)*: The extent to which the preconditions (α), postconditions (β), predecessors (λ), and contributing goals (μ) of a goal are SATISFIED or UNSATISFIED makes beliefs about the goal stronger. An UNKNOWN pre/postcondition status of a goal and its predecessors and contributing goals forms weaker beliefs. For instance, if one knows all predecessors of a pursued goal (e.g., "Press Bread Slices") are SATISFIED (i.e., "Spread Peanut Butter" and "Spread Strawberry Jam"), failure of the pursued goal will elicit one's negative emotion (due to the strong beliefs related to the goal); whereas not knowing the status of the goal-related factors (e.g., whether Mary could find the knife to cut the sandwich in half) causes one to form weaker beliefs about the goal.
- *Saliency (S)*: 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. For example, according to Figure 2, if one is making a peanut butter sandwich, all of the *live* (see Section 4) contributing goals to "Prepare Peanut Butter Sandwich" (e.g. Spread Peanut Butter) will be in the focus stack, and consequently are more salient. *Non-live* goals in the shared plan (e.g. Slice Egg) will not be salient.
- *Persistence (P)*: The recurrence of a belief over time (turns) increases the persistence of the belief. Beliefs occurring only once have the lowest value of persistence. For instance, if Mary keeps saying that she can not find the knife to cut the sandwich in half, one could pursue a new goal outside of the shared plan to acknowledge Mary's concern.

We also use two motive attributes discussed in Section 3.1 to compute the motive related part of the utility ($\mathcal{U}$):

- *Urgency (γ)*: 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. For instance, if one has a private goal to prepare for making the second sandwich (e.g. get the eggs) while Mary is waiting to get the first sandwich and cut it in half, pressing bread slices and passing them to Mary will be more urgent than one's private goal.
- *Importance (η)*: 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. For example, if one cannot find white bread on which to spread peanut butter (an impasse to make the peanut butter sandwich), and Mary offers to use wheat bread instead (external motive), the new motive becomes important to remove the impasse in the shared plan.

$$\Psi = \frac{\alpha_k + \beta_k + \lambda_k + \mu_k}{\alpha_{all} + \beta_{all} + \lambda_{all} + \mu_{all}} + \eta + \gamma \qquad (1)$$

$$\begin{array}{rl} \eta, \gamma \in \mathbb{N}, & \eta, \gamma \geq 0 \\ \alpha_k, \beta_k, \lambda_k, \mu_k \in \mathbb{N}, & \alpha_k, \beta_k, \lambda_k, \mu_k \geq 0 \\ \alpha_{all}, \lambda_{all}, \mu_{all} \in \mathbb{N}, & \alpha_{all}, \lambda_{all}, \mu_{all} \geq 0 \\ \beta_{all} \in \mathbb{N}, & \beta_{all} \geq 1 \end{array}$$

In equation 1, the subscript *k* refers to the *known* goal-related factors (SATISFIED or UNSATISFIED); whereas the subscript *all* includes both *known* and *unknown* goal-related factors. In this equation, both urgency (γ) and importance (η) attributes of motives can impact the outcome of the goal-related belief attributes' ratio, and ultimately the Ψ value.

$$U(\varepsilon_t) = \begin{cases} \upsilon P \cdot S^{\Psi} & \Psi > 0 \\ 0 & \Psi = 0 \end{cases} \qquad (2)$$

We compute the utility of an event based on the above five attributes. The υ's value becomes +1 if the status of the corresponding goal is ACHIEVED, PENDING, or IN PROGRESS, and υ's value becomes -1 if the status of the corresponding goal is FAILED, BLOCKED, INAPPLICABLE. Therefore, the value of the overall utility can be positive, negative or zero.

The significance of an event in a collaborative environment is based on the utility of the event and the perceived emotion of the human collaborator. The human's emotion influences the relevance 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 makes the robot find an event relevant with even a slightly positive utility. This threshold value is currently determined based on whether the valence of the human's perceived emotion is positive (e.g., happiness) or negative (e.g., anger). Therefore, an event can be considered IRRELEVANT even though the utility has a relatively positive value, because relevance is influenced by the human's perceived negative emotion.

5.2 Controllability

Controllability is the extent to which an event can be influenced, and it is associated with a robot's ability to cope with 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 2 (Controllability)

1: **function** IsEventControllable(Event ε_t)
2: $\quad g_t \leftarrow$ *recognizeGoal*(ε_t)

3: $\quad \mathcal{M} \leftarrow$ GetAgencyRatio(g_t)
4: $\quad \mathcal{R} \leftarrow$ GetAutonomyRatio(g_t)
5: $\quad \mathcal{P} \leftarrow$ GetSucPredecessorsRatio(g_t)
6: $\quad \mathcal{I} \leftarrow$ GetAvailableInput(g_t)
7: $\quad \mathcal{V}_{e_h} \leftarrow$ getEmotionValence(g_t)
8: $\quad \omega \leftarrow$ getWeights(g_t)

9: $\quad \mathcal{X} \leftarrow \frac{\omega_0 \cdot \mathcal{M} + \omega_1 \cdot \mathcal{R} + \omega_2 \cdot \mathcal{P} + \omega_3 \cdot \mathcal{I}}{\omega_0 + \omega_1 + \omega_2 + \omega_3} + \mathcal{V}_{e_h}$

10: $\quad$ **if** $(\mathcal{X} > 0)$ **then**
11: $\quad\quad$ **return** CONTROLLABLE
12: $\quad$ **else**
13: $\quad\quad$ **return** UNCONTROLLABLE

Controllability is 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 2, we compute the controllability of an event based on the above four factors (lines 3 to 6). We use weighted averaging over these four factors to compute the aggregated value of the controllability of an 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. We believe that the human's perceived emotion can also impact the aggregated value of controllability of an event ($\mathcal{V}_{e_h}$ in line 9). The $(-1.0 \leq \mathcal{V}_{e_h} \leq 1.0)$ is the valenced value of the human's perceived emotion. Positive emotion, e.g., happiness, possess positive valence value, and negative emotions, e.g., anger, have negative value of valence. The magnitude of this value can change with respect to the intensity of the perceived emotion. Furthermore, the overall aggregated value indicates whether the current event is controllable, if positive; otherwise the event will be considered as uncontrollable.

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 $\mathcal{M}$ in Algorithm 2 (i.e., $\mathcal{M} = 1.0$) if the robot's mental state possesses an internal motive towards the recognized goal; otherwise we consider the minimum agency value (i.e., $\mathcal{M} = 0.0$) 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 2, $\mathcal{R}$ 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 $(0 \leq \mathcal{R} \leq 1)$. However, if the associated goal of the current event corresponds to a primitive goal the value of $\mathcal{M}$ would be 0.0 or 1.0. 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_t, are other goals that the collaborators should achieve before trying to achieve goal g_t. 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 $\mathcal{P}$ in Algorithm 2) 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 $\mathcal{P} = 1.0$ which is the maximum value for $\mathcal{P}$. On the contrary, failure of all of the predecessors will lead to $\mathcal{P} = 0.0$. Therefore, a higher $\mathcal{P}$ 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 $\mathcal{I}$ in Algorithm 2) will be bound to 0.0 and 1.0. Similar to the other factors in the controllability process, the closer the value of $\mathcal{P}$ gets to 1.0, the more positive impact it has on the overall controllability value of the event.

In summary, the output of these two and other appraisal processes such as *desirability* [Shayganfar *et al.*, 2016a] and *expectedness* [Shayganfar *et al.*, 2016c] serves as critical input for the other mechanisms and processes (e.g., goal management [Shayganfar *et al.*, 2016b]) 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.

6 Evaluation and Discussion

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 two 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 each appraisal variable. There were 12 questions (including 2 test questions) in the controllability, 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 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
Controllability	33	0.743	0.158	<0.001

To minimize the background knowledge necessary for our test subjects, we used a simple 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 both 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 2 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. Note that the collaboration structure and the instructions were the same for both questionnaires.

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

Each question was designed based on different factors that we use in our algorithms (see Section 5). Here, we present one example question from the relevance questionnaire, and describe how this question relates to a specific factor within the corresponding algorithm. The input for our algorithms was the task model depicted in Figure 2.

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 3: Example Expectedness Question.

Figure 3 shows an example question from the relevance questionnaire. In this example, with respect to Algorithm 1 (line 6), option A is relevant because the task related to this option provides the next available task in the focus stack (see the task model in Figure 2). Although the task in option B is part of the existing task model, it is considered as irrelevant 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 relevance of this event is similar to the output of the relevance algorithm. For this question, the human decision was 97% similar to the algorithm's output. Average results for both questionnaires are presented in Table 1.

We conducted the user study to compare the results with the implemented algorithms discussed in Section 5. 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 agreement ratio 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 *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 judgments of humans 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 two appraisal algorithms predicted the human judgments.

7 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

Acknowledgment is suppressed for anonymity.

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.

[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.*, 2016a] Mahni Shayganfar, Charles Rich, and Candace Sidner. Toward improving human-robot collaboration with emotional awareness. *Under Review*, 2016.

[Shayganfar *et al.*, 2016b] Mahni Shayganfar, Charles Rich, and Candace L. Sidner. Impact of affective appraisal on collaborative goal management: My robot shares my worries. In *Proceedings of 11th ACM-IEEE International Conference on Human-Robot Interaction (LBR)*, 2016.

[Shayganfar *et al.*, 2016c] 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.

[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.