RESEARCH STATEMENT

Mahni Shayganfar (mshayganfar@wpi.edu)

My current research span the areas of computational collaboration theories, affective computing, human-robot collaboraiton, and cognitive robotics. A common thread in my research is in developing a theory (Affective Motivational Collaboration Theory), design of a domain-independent architecture, and the framework which uses this architecture to provide a collabortive behavior for robots or vritual agents. I have resorted to prominent computational collaboration theories, i.e., SharedPlans theory, and computaitonal models of emotions, i.e., cognitive appraisal theory to develop my own theory. Broadly speaking, my research belongs to the area of human-robot collaboration and its underlyign processes, a growing field which has influence on different leading industries such as autonomous vehicles, space exploration, manufacturing, and any industry including situation inwhich human-robot teamwork is required.

Background - Collaboration Theories

The construction of robots that are intelligent, collaborative problem-solving partners is important in robotics and applications of Artificial Intelligence. It has always been important for us to make robots better at helping us to do whatever they are designed for. To build collaborative robots, we need to identify the capabilities that must be added to them so that they can work with us or other agents. As Grosz says, collaboration must be designed into systems from the start; it cannot be patched on [12]. Collaboration is a special type of coordinated activity in which the participants work together performing a task or carrying out the activities needed to satisfy a shared goal [15].

Collaboration involves several key properties both in structural and functional levels. For instance, most collaborative situations involve participants who have different beliefs and capabilities; most of the time collaborators only have partial knowledge of the process of accomplishing the collaborative activities; collaborative plans are more than the sum of individual plans; collaborators are required to maintain mutual beliefs about their shared goal throughout the collaboration; they need to be able to communicate with others effectively; they need to commit to the group activities and to their role in it; collaborators need to commit to the success of others; they need to reconcile between commitments to the existing collaboration and their other activities; and they need to interpret others' actions and utterances in the collaboration context [13]. These collaboration properties are captured by the existing computational collaboration theories.

As I mentioned, to be collaborative, partners, e.g., a robot and a human, need to meet the specifications stipulated by collaboration theories. These theories argue for an essential distinction between a collaboration and a simple interaction or even a coordination in terms of commitments [11, 23]. The prominent collaboration theories are mostly based on plans and joint intentions [4, 17, 22], and they were derived from the BDI paradigm developed by Bratman [2] which is fundamentally reliant on folk psychology [27]. The two theories, Joint Intentions [4] and SharedPlans [17], have been extensively used to examine and describe teamwork and collaboration.

SharedPlans theory - The SharedPlans model of collaborative action, presented by Grosz and Sidner [14, 15, 17], aims to provide the theoretical foundations needed for building collaborative robots/agents [12]. SharedPlans is a general theory of collaborative planning that requires no notion of joint intentions, accommodates multi-level action decomposition hierarchies and allows the process of expanding and elaborating partial plans into full plans. SharedPlans theory explains how a group of agents can incrementally form and execute a shared plan that then guides and coordinates their activity towards the accomplishment of a shared goal. SharedPlans is rooted in the observation that collaborative plans are not simply a collection of individual plans, but rather a tight interleaving of mutual beliefs and intentions of different team members.

Joint Intentions theory - Following Bratman's guidelines, Cohen and Levesque propose a formal approach to building artificial collaborative agents. The Joint Intentions theory of Cohen and Levesque [4, 5, 6, 7, 21] represents one of the first attempts to establish a formal theory of collaboration, and due to its clarity and expression, is a widely used teamwork theory. The basic idea of Joint Intentions theory is based on individual and joint intentions (as well as commitments) to act as a team member. Their notion of joint intention is viewed not only as a persistent commitment of the team to a shared goal, but also implies a commitment on part of all its members to a mutual belief about the state of the goal. In other words, Joint Intentions theory describes how a team of agents can jointly act together by sharing mental states about their actions while an intention is viewed as a commitment to perform an action. A joint intention is a shared commitment to perform an action while in a group mental state [5].

STEAM - Tambe in [34] argues that teamwork in complex, dynamic, multi-agent domains requires the agents to obtain flexibility and reusability by using integrated capabilities. Tambe created STEAM (simply, a **S**hell **TEAM**work) based on this idea. STEAM's operationalization in complex, real-world domains is the key in its development to addressing important teamwork issues. STEAM is founded on the Joint Intentions theory and it uses joint intentions as the basic building block of teamwork while it is informed by key concepts from SharedPlans theory.

I believe the SharedPlans and Joint Intentions collaboration theories are the most well-defined and well-established theories in computer science. I found SharedPlans theory more convincing than the other major and subordinate approaches, with respect to its inclusive explanation of the collaboration structure and its association to discourse analysis which directly improves the communicative aspects of a collaboration theory. I also understand the value of Joint Intentions theory due to its clarity and closeness to the foundations of collaboration concepts. These specifications of the Joint Intentions theory can make it applicable in multi-agent system designs and human-robot collaboration. I also consider hybrid approaches valuable, such as STEAM, if they clearly understand drawbacks with existing theories and successfully achieve better collaborative agents by infusing different concepts from different theories.

Background - Cognitive Appraisal Theory

According to Picard [26], the term affective computing encapsulates a new approach in Artificial Intelligence, to build computers that show human affection. Studies show that the decision making of humans is not always logical [10], and in fact, not only is pure logic not enough to model human intelligence, but it also shows failures when applied in artificial intelligence systems [8]. Emotions impact fundamental parts of cognition including perception, memory, attention and reasoning [3]. This impact is caused by the information emotions carry about the environment and event values. If we want robots and virtual agents to be more believable and efficient partners for humans, we must consider the personal and social functionalities and characteristics of emotions; this will enable our robots to coexist with humans, who are emotional beings.

Cognitive Appraisal Theory - There are different types of computational theories of emotion. These theories differ in the type of relationships between their components and whether a particular component plays a crucial role in an individual emotion. Appraisal theories of emotion were first formulated by Arnold [1] and Lazarus [20] and then were actively developed in the early 80s by Ellsworth and Scherer and their students [28, 29, 30, 32, 33]. The emotional experience is the experience of a particular situation [9]. Appraisal theory describes the cognitive process by which an individual evaluates the situation in the environment with respect to the individual's well-being and triggers emotions to control internal changes and external actions. According to this theory, appraisals are separable antecedents of emotion, that is, the individual first evaluates the environment and then feels an appropriate emotion [33]. The appraisal procedure begins with the evaluation of the environment according to the internalized goals and is based on systematic assessment of several elements [31]. The outcome of this process triggers the appropriate emotions.

In my Ph.D thesis, I attempt to lay a computational framework for the theory I have developed based on SharedPlans and Cognitive Appraisal theories. A great deal of my work has benefited from the integration of these well-established theories and their underlying structure.

Limitations

There are several well-developed cognitive architectures, e.g., Soar [19] and ACT-R [18], each with different approaches to defining the basic cognitive and perceptual operations. There have also been efforts to integrate affect into these architectures [24]. In general, however, these cognitive architectures do not focus on processes to specifically produce affect-regulated goal-driven collaborative behaviors. At the same time, existing collaboration theories, e.g. SharedPlans theory [17], Joint Intentions [4], and STEAM [34], focus on describing the structure of a collaboration in terms of fundamental mental states, e.g., mutual beliefs or intentions. Although all the existing collaboration theories are well-defined and properly introduce collaboration concepts, they mostly explain the structure of a collaboration and they lack the underlying domain-independent processes with which collaborative procedures could be defined more systematically and effectively in different applications. Therefore, it is crucial to investigate the cognitive processes involved in a collaboration in the context of a cognitive architecture to be able to describe the underlying processes, their relationships, and their influences on each other.

I believe that the evaluative role of emotions as a part of cognitive processes helps an agent to perform appropriate behaviors during a collaboration. It is important to think about the underlying cognitive processes of the collaborators in order to have a better understanding of the role of emotions. To work jointly in a coordinated activity, participants (collaborators) act based on their own understanding of the world and the anticipated mental states of the counterpart; this understanding is reflected in their collaborative behaviors. Emotions are pivotal in the collaboration context, since their regulatory and motivative roles enhance an individual's autonomy and adaptation as well as his/her coordination and communication competencies in a dynamic, uncertain and resource-limited environment. The collaborative behavior of the individuals can also be influenced by the tasks contributing towards a shared goal. Some tasks may be inherently insignificant, boring, unpleasant or arduous for a collaborator. Thus, knowing how to externally motivate other collaborator to perform such tasks becomes an essential skill for a participant in a successful collaboration. Such knowledge enables an individual to lead his collaborator to internalize the responsibility and sense of value for an externally motivated task.

Current Research

In my Ph.D. thesis, I have developed the Affective Motivational Collaboration Theory and the associated computational model that will enhance the performance and effectiveness of collaboration between robots and humans. This theory explains the functions of emotions in a dyadic collaboration and shows how affective mechanisms can coordinate social interactions by anticipating other's emotions, beliefs and intentions. This theory also specifies the influence of the underlying collaboration processes on appraisals. Affective Motivational Collaboration Theory elucidates the role of motives as goal-driven emotion-regulated constructs with which an agent can form new beliefs and intentions to cope with internal and external events. An important contribution of this work is to elucidate how motives are involved not only in the appraisal and coping processes, but how they also serve as a bridge between appraisal processes and the collaboration structure. I will validate my theory using my computational framework in the context of a human-robot collaboration.

I have investigated how affect-driven mechanisms in my Affective Motivational Collaboration Theory can be involved to successfully maintain a dyadic collaboration. This study is based on two hypothetical collaboration examples of *task delegation* and *agreeing on a shared goal* for both positive and negative situations. In positive cases, two walkthrough examples are provided to show successful progress of a collaboration in each scenario. In these examples collaboration is concluded by the shared goal achievement because of the emotion-aware behavior of a robot. Whereas, the negative cases, illustrate how emotion-ignorant behavior of a robot can cause failure of a collaboration in the same examples. We have implemented the rules associated with these examples using JESS (Java Expert System Shell) which is a rule engine for the Java platform. In our current implementation we have categorized the rules in different modules associated with the mechanisms and the underlying processes in Affective Motivational Collaboration Thoery. We are implementing algorithms for each individual mechanism to be able to automatically generate the required facts within each mechanism to fire the existing rules. Ultimately, we are going to offer a platform which operates based on the collaboration structure discussed by SharedPlans theory [16], and employs affect-driven processes such as the appraisal processes in [25] to enable a robot to obtain and demonstrate collaborative behaviors. This study is currently under review of a journal publication.

As another crucial part of my research, I have developed domain-independent algorithms for the appraisal of a collaborative environment. These algorithms compute the value of the four appraisal variables, i.e., *relevance, desirability, expectedness* and *controllability.* These algorithms use a generalized collaboration structure to compute their outcome in a dyadic collaboration. The performance of individual algorithms will be verified based on the analysis of the answers human subjects provide to our questionnaire. All of the questions have been prepared based on the same task models provided to our framework. I will finalize my Ph.D. thesis by evaluating the overall architecture using a robot simulator to successfully and efficiently (in terms of time and number of errors) collaborate with human subjects.

A Research Agenda

Collaborative robots are becoming an integral part of humans' environment to accomplish their industrial and household tasks. In these environments humans are involved in robots' operations and decision making processes. This involvement of humans influences the efficiency of robots' interaction and performance, and makes them dependent on the humans' cognitive abilities and mental states. Similarly, humans' performance can be maintained using robot's adaptive behavior. The evaluative nature of affective processes can assist a robot to perform adaptive behaviors with respect to the current goal. Goal management, alarm mechanism and action selection are some examples of the functions that a system with affect-driven processes offers.

Performance evaluation of a human requires perception of the human's physical and mental states. This perception can be performed based on available bio-information, as well as verbal (e.g., utterancces) and non-verbal (e.g., facial expressiona) behaviors. However, a robot requires to possess cognitive mechanisms to evaluate and adapt its own behavior to change human's performance on the given task and environment. My Ph.D. thesis provides fundamental affect-driven mechanisms and defines their relationship within a cognitive architecture. These mechanisms can be involved form initial evaluation of the input data to higher level processes such as action selection or alarm mechanism. All these functions can improve robot's awareness and self-synchronization capabilities during the interaction. For instance, if the appraisal of the input data reveals controllability (an appraisal varibale) of an event, the robot can choose an appropriate action to make changes to the status of the interaction. Ultimately, these changes can improve human's performance in a task execution. I would like to apply my knowledge I attained in my research and develop new algorithms to evaluate performance of a human and adapt the robot's behavior respectively.

References

[1] Magda B. Arnold. *Emotion and personality.* Cassell Co., 1960.

[2] Michael E. Bratman. *Intention, Plans, and Practical Reason.* Cambridge, Mass.: Harvard University Press, 1987.

[3] Gerald L. Clore and Jeffrey R. Huntsinger. How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9):393–399, 2007.

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

[5] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261, 1990.

[6] Philip R. Cohen and Hector J. Levesque. Persistence, intention, and commitment. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 33–69. MIT Press, Cambridge, MA, 1990.

[7] Philip R. Cohen, Jerry Morgan, and Martha E. Pollack. *Intentions in Communication*. A Bradford Book, 1990.

[8] Hubert L. Dreyfus. *What Computers Still Can'T Do: A Critique of Artificial Reason*. MIT Press, 1992.

[9] Nico H. Frijda. *The Emotions*. Cambridge University Press, 1986.

[10] S. Grossberg and W. E. Gutowski. Neural dynamics of decision making under risk: Affective balance and cognitive-emotional interactions. *Psychological Review*, 94(3):300–318, 1987.

[11] Barbara Grosz and Sarit Kraus. The evolution of shared plans. In *Foundations and Theories of Rational Agency*, pages 227–262, 1998.

[12] Barbara J. Grosz. AAAI-94 presidential address: Collaborative systems. *AI Magazine*, 17(2):67–85, 1996.

[13] Barbara J. Grosz. Beyond mice and menus. *Proceedings of the American Philosophical Society*, 149(4):523–543, 2005.

[14] Barbara J. Grosz, Luke Hunsberger, and Sarit Kraus. Planning and acting together. *AI Magazine*, 20(4):23–34, 1999.

[15] Barbara J. Grosz and Sarit Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[16] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, July 1986.

[17] 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.

[18] Christian Lebiere John Robert Anderson. *The Atomic Components of Thought*. Lawrence Erlbaum Associates, 1998.

[19] John Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.

[20] Richard Stanley Lazarus. *Emotion and Adaptation*. OXFORD University Press, 1991.

[21] Hector J. Levesque, Philip R. Cohen, and Jos H. T. Nunes. On acting together. In *AAAI*, pages 94–99. AAAI Press / The MIT Press, 1990.

[22] D. J. Litman and J. F. Allen. Discourse processing and commonsense plans. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 365–388. MIT Press, Cambridge, MA, 1990.

[23] Karen E Lochbaum. A collaborative planning model of intentional structure. *Computational Linguistics*, 24(4):525–572, 1998.

[24] Robert P. Marinier III, John E. Laird, and Richard L. Lewis. A computational unification of cognitive behavior and emotion. *Cognitive System Research*, 10(1):48–69, March 2009.

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

[26] Rosalind W. Picard. *Affective Computing*. The MIT Press, 2000.

[27] Ian Ravenscroft. *Folk Psychology as a Theory*. Stanford Encyclopedia of Philosophy, 2004.

[28] Ira J. Roseman and Craig A. Smith. Appraisal theory:overview, assumptions, varieties, controversies. In Klaus R. Scherer, Angela Schorr, and Tom Johnstone, editors, *Appraisal process in emotion*, pages 3–34. NY: Oxford University Press, 2001.

[29] David Sander, Didier Grandjean, and Klaus R. Scherer. A systems approach to appraisal mechanisms in emotion. *Neural Networks*, 18(4):317–352, 2005.

[30] Klaus R. Scherer. On the nature and function of emotion: A component process approach. In Klaus R. Scherer and Paul Ekman, editors, *Approaches To Emotion*, pages 293–317. Lawrence Erlbaum, Hillsdale, NJ, 1984.

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

[32] Klaus R. Scherer. Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3459–3474, 2009.

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

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