The Social Agency Problem

Daniel Shapiro
Institute for the Study of Learning and Expertise
2164 Staunton Court, Palo Alto, CA 94036
dgs@isle.org

Abstract

This paper proposes a novel agenda for cognitive systems research focused on the "social agency" problem, which concerns acting to produce mental states in other agents in addition to physical states of the world. The capacity for social agency will enable agents to perform a wide array of tasks in close association with people, and is a valuable first step towards broader social cognition. We argue that existing cognitive systems have not addressed social agency because they lack a number of the required mechanisms. We describe an initial approach set in toy scenario, based on capabilities native to the Icarus cognitive architecture. We utilize an analysis of this approach to highlight the open issues required for social agency, and to encourage other researchers to address this important problem.

Introduction and Motivation

Social cognition is a critically important aspect of human intelligence that touches all aspects of our lives. It spans our abilities to recognize the intentions of people from their behavior, reason about the cares and concerns of our children to make them happy, collaborate to solve business tasks, and to act both cooperatively and competitively in team sports while anticipating the behavior of others. Social cognition lets us frame and evaluate decisions ranging from moving to marriage by considering the effects on ourselves, and others. It also helps us to extract meaning from our experiences by contemplating the value we gain and provide via our interactions. Social cognition is so central to behavior that many ethologists believe it drove the evolution of the human brain (Byrne and Whiten, 1988). If computational systems are ever going to achieve human level intelligence, we must address social cognition.

Cognitive systems, by comparison, focus far more on world state than achieving social goals. Game playing systems consider the intentions of opponents, but only indirectly as a part of searching board positions (Hsu, 2002). Cognitive prostheses augment human intelligence, but typically by offloading component functions. For example, Garvey, et al. (2011) automates elements of military plan generation after learning how to perform routine tasks. Virtual human characters supply intensely interactive experiences, but their designs minimize knowledge about others. For example, Swartout (2010) develops conversational characters for military training that react to user queries by supplying information. These illustrations suggest that significant cognitive systems lack the capacity for social interaction that characterizes human collaboration.

We believe that the increasing capability of cognitive systems provides an opportunity to go beyond these bounds. In particular, we propose that future research should recast cognitive systems as social actors vs. reasoning tools as a means of driving new technology and enabling new applications.

The long-term goal of this research will be to support broad social cognition. However, that objective may be impractical in the near term, as it requires an implemented understanding of social roles and schema, their unfolding over time, as well as the ability to recognize state relative to those constructs from observation. We lack the relevant computational models, the sensory bandwidth, and the interpretation functions at this time. However, a first step towards social cognition is within reach now, and it will yield a significant and useful capacity for behavior in social spheres.

We advocate a research agenda to produce what we call “social agency”. Cognitive systems of this form will have the capacity to model the mental state of others, hold intentions towards others, and choose behavior cognizant of the likely actions and reactions of others. We can also state this problem in more precise terms.

Given:

- a model of the observations, available inferences, beliefs, skills, goals, and preferences of others,
- an agent’s own observations, available inferences, beliefs, skills, and preferences,

- a desired belief state for self and others, and
- a desired physical state of the world,

Generate:
- a sequence of actions that produces the desired belief and world state in self and others, given the responses and independent actions of other agents.

We claim that a capacity for social agency will enable a number of beneficial applications. For example, automated counselors and coaches should be fundamentally concerned with a user's mental state. Automated patients, adversaries, and non-player characters should benefit from consideration of what the user knows or is thinking. More generally, an appreciation of other agents' knowledge and skills will let cognitive systems participate in task decomposition and delegation decisions, supporting more sophisticated forms of collaboration. Finally, cognitive systems with social agency can plausibly perform better on existing tasks (like military planning and training) through attention to the user mental states that motivate, and/or are affected by system behavior.

The remainder of this paper explores various facets of the social agency problem. We intend this as an exploratory research report, so we follow advice due to Dietterich (1990) by describing, in order, the challenges of this endeavor, the limitations of existing approaches, and an initial approach (based on the current capabilities of the Icarus cognitive architecture). Then, we use an analysis of the limitations of this approach to propose a research agenda on social agency. We end with concluding remarks.

Challenges of Social Agency

We have claimed that the social agency problem is novel because of its strong focus on the mental state of other agents. This expanded information base impacts virtually all of the cognitive machinery that generates performance from knowledge. As a result, social agency gives rise to a number of research challenges. These concern the content and origin of mental models, as well as their use in behavior generation.

A first challenge is to identify and represent the distinctions necessary to model the mental state of others. At minimum, social agency requires constructs reflecting what agent A believes agent B knows, cares about, or can do, as well as statements about the goals, preferences, intent, and motivations imputed to other agents. The specific vocabulary for these distinctions is an open question. However, given that the goal is to reason about the knowledge and behavior of other agents, especially human collaborators, we can draw on psychological, emotional, and decision theoretic frameworks to define structures that can be shared across multiple user models.

The emphasis on mental models also implies a challenge to assess and verify their content from observation. Given that distinctions about mental state are unobservable, the data is likely to be both evidential and sparse. As result, the assessment task has an explanatory and/or an abductive nature.

Social agency also presents a challenge to represent and reason about complex intent. This includes assertions that agent A wants agent B to believe x, intend y, or hold goal z, as well as to execute an action, possess a skill, prefer one state over another, value a condition, and be in a given motivational state. These forms occur naturally in human collaboration but fall largely outside the territory explored by cognitive systems researchers.

A related challenge concerns the origin of complex intentions. Research on social agency will either need to impose them by design, or derive them via problem solving. The design perspective asks what one agent *should* intend towards another in order to accomplish a joint task. This can be addressed in a problem specific fashion, while a more general understanding might call for establishing a correspondence in motivational states. The approach of deriving complex intentions emphasizes issues of goal creation, selection, and rejection. Recent research on goal directed autonomy can inform this work (Choi, 2010; Klenk, 2010).

Finally, social agency poses a significant challenge to utilize knowledge about the mental states and intentions of other agents for behavior generation. This information places new demands on the machinery for reasoning, planning and problem solving, and execution. For example, inference needs to assess mental state, planning must consider the motivations and likely actions of other agents, and execution must respond to the new sources of uncertainty in this environment. The need to capture motivational structures deserves special mention, since the goal is to predict the behavior of human participants.

In summary, social agency requires a cognitive system to model the mental structures of agents and people. This imposes significant research challenges. At the same time, insight into the mental state of human users promises new leverage for intelligent agent behavior.

Limitations of Current Approaches

In addition to defining a novel problem, an exploratory research paper should clarify why existing methods are insufficient to solve it. We consider cognitive systems that currently operate in social realms, methods for reasoning about the mental states of other agents, and general theories of cognition, below.

Conversational virtual humans are designed to provide engaging social interactions surrounding an application task. They have been applied to develop synthetic characters for training tasks in military mission rehearsal, negotiation management, and PTSD counseling (Swartout, 2010), as well as museum guides (Leuski and Traum, 2011). While these characters generate behavior via complex cognitive processing (including reasoning, planning, and emotional dynamics), they possess very shallow user models. They react to single user utterances taken in isolation as a means of reducing the complexity of the application task. Similarly, Wray et al. (2004) employ the SOAR cognitive architecture to create intelligent adversaries for games, which reason, make plans and project likely opponent behavior, although they focus on achieving world state without considering the opponent's mental state. In contrast, intelligent tutoring systems maintain strong user models, which they employ, for example, to recognize knowledge state, diagnose errors, and deliver appropriate curricula (Greer, 1994). These systems tend to react to student knowledge states vs. plan to achieve them, which is important for social agency.

Research on plan understanding centrally concerns the mental state of others. For example, Wilensky (1983) considers story fragments and infers that an agent holds plans to explain observed actions, and goals to explain plans. These plans and goals come from a rich vocabulary of domain specific and stereotypical behavior templates that has its roots in conceptual dependency networks. This style of analysis addresses relevant abduction tasks, but the templates are unlikely to generalize to the forms of complex intent found in social agency.

Research on modal logics has resulted in a large family of formalisms that address representation and reasoning tasks related to social agency. These include logics of knowledge and belief across agents, collaboration and commitment, and the manipulation of beliefs, desires, and intentions (Woolridge, 2000). These systems have been applied to generate behavior in practical contexts (Rao and Georgeff, 1995). However, strong questions remain about their sufficiency, practicality, and convenience for use in social cognition, especially to address the required capabilities in combination.

Systems that articulate the cognitive processes within single agents in general form are highly relevant to social agency. General theories of cognition are the best examples, such as SOAR (Laird, et al, 1987), Icarus (Langley, et al 2009) and Act-R (Anderson and Lebiere, 1998). These systems provide multiple cognitive mechanisms that perform inference, planning, problem solving, and execution. Motivational constructs have also been incorporated into these architectures. For example, work by Gratch and Marcella (2001) employs emotion in Soar to guide planning and understanding. Work on value-driven agents in Icarus (Shapiro, 2001) employs value models to drive decision-making and learning. However, none of these systems provides constructs for modeling the cognitive functions of other agents.

In summary, current approaches provide a variety of tools for addressing social agency, but lack a framework that combines multiple cognitive functions with the capacity to reason about others.

Social Agency in Icarus

Now that we have defined the social agency problem and clarified its relation to current technology, we describe an initial approach to a subset of the challenges it presents for cognitive systems research.

We claim that it is natural to investigate social agency in the context of cognitive architectures, as they are designed to manipulate many of the mental distinctions that people possess. We have chosen to implement our initial approach within the Icarus framework (Langley, et al, 2009), which is an instance of a general theory of cognition. We utilize Icarus' existing representations and reasoning methods with minor extension, in part to show that current technology affords feasible first steps towards social cognition, and in part to illustrate the limitations of existing methods en route to clarifying a research agenda.

Simple Sibling Interaction

We begin by formulating a minimal problem that requires social agency. We phrase this problem in TWIG (Horswill, 2008), which is a simulated environment designed to foster agent research. It provides multiple agents, a small family of object types, plus a low-resolution graphical interface.

Figure 1 illustrates a sibling interaction task. It involves two agents, nominally a brother and sister, plus a doll and a ball, where the brother is "self" and the sister is the "other" agent in the scenario. Stated informally, the brother has the problem to ensure that both he and his sister are in the 'comforted' state, where playing with toys generally denotes comfort. However, the brother also believes that the sister will become jealous, and not comforted, if he is playing with her favorite toy.

The initial state has the sister holding her favorite toy (the doll), and the brother empty handed. As a result, she is

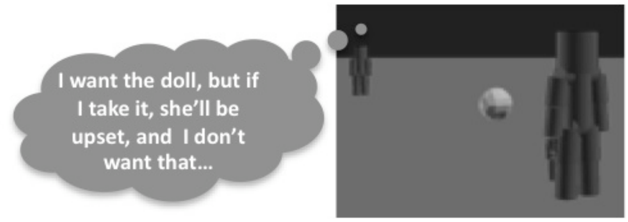

Figure 1. A sibling interaction scenario.

comforted but he is not. The brother has skills to take the doll or to find another toy, so his problem is to chart the right path given his model of his sister's mental state.

Icarus Problem Representation

We can state the sibling interaction problem in formal terms using representations native to the Icarus architecture. Icarus expresses concept and skill knowledge in hierarchical relational structures, beliefs as relational instances, and problems as conjunctive goals.

Table 1 illustrates Icarus concepts for this scenario. These should be understood as definitions for mental states, expressed as conjuncts of relations containing unbound variables. Here, the brother agent is comforted if it has a toy. The sister is jealous when her favorite toy comforts the brother. Has-toy is defined slightly differently depending upon the object; the agent must be near a ball, or holding a doll. This hierarchy grounds out in percepts obtained from the environment, in this case, objects representing the agents, the doll, and the ball.

Note that these concepts are intended to capture the brother's understanding of the world. That is, the brother attributes mental states to the sister, called comforted and jealous, and assumes a relation between them that also depends upon his own state.

Table 1. Icarus concepts for sibling interaction.

```
((comforted ?agent ?toy)
  :relations ((brother ?agent)
              (has-toy ?agent ?toy)))

((comforted ?agent ?toy)
  :relations ((sister ?agent)
              (has-toy ?agent ?toy)))
              (not (jealous ?agent ?toy2))))

 ((jealous ?sister ?toy)
  :relations ((sister ?sister)
              (brother ?brother)
              (comforted ?brother ?toy)
              (favorite-toy ?sister ?toy)))

((has-toy ?agent ?toy)
  :percepts ((twig.child ?agent) (twig.ball ?toy))
  :relations ((at-ball ?agent ?toy) (toy ?toy)))

((has-toy ?agent ?toy)
  :percepts ((twig.child ?agent) (twig.doll ?toy))
  :relations ((holding ?agent ?toy) (toy ?toy)))

;; The sister's favorite toy is a doll
```

Table 2 illustrates two of the brother's skills for the sibling interaction scenario, to take the doll, and to get a ball. Icarus skills are hierarchical relational structures with required conditions, subskills (steps to take) for completing the task, and expected effects. For example, the brother can only take the doll if the sister has it. In order to physically take it, he must move to the sister, cause her to drop the doll, and pick it up for himself, resulting in a situation where he is holding the doll and she is not.

Table 2. Icarus skills for sibling interaction.

```
((take-doll ?brother ?doll)
  :conditions ((brother ?brother)
               (sister ?sister)
               (has-toy ?sister ?doll)
               (doll ?doll))
  :subskills ((move ?brother ?doll)
              (drop ?sister ?doll)
              (lift ?brother ?doll))
  :effects ((holding ?brother ?doll))
            (not (holding ?sister ?doll))))

 ((get-ball ?brother ?ball)
  :conditions ((ball ?ball)
               (brother ?brother)
               (not (at-ball ?any ?ball)))
  :subskills ((move ?brother ?ball))
  :effects ((at-ball ?brother)))
```

Given these structures, we can complete the remaining elements of the problem definition by identifying the initial state, and the brother's desired physical and mental states for the sibling interaction scenario.

Beliefs in Icarus are literals corresponding to concepts whose definitions hold true in the current environment. The initial state contains several beliefs representing background knowledge: (sister Sister1), (brother Brother1), (doll doll1), (ball ball1), (toy doll1) (toy ball1) (favorite-toy Sister1 doll1). It also contains beliefs inferred from current percepts: (holding Sister1 Doll1) (has-toy Sister1 Doll1), and (comforted Sister1 Doll1).

Given these structures, the brother's problem statement is quite simple. He must generate a sequence of actions that produces the conjunction of two mental states:

((Comforted Brother1 ?toy1) (Comforted Sister1 ?toy2))

Icarus Performance Elements

Icarus solves this problem through use of three performance elements; conceptual inference, skill execution, and problem solving. Since this paper primarily concerns challenge problems, we will only provide an overview here.

Inference in Icarus produces belief instances, given current percepts and the conceptual hierarchy as inputs. On each cycle, the environment deposits descriptions of perceived objects into a perceptual buffer. Inference compares these precepts against conceptual rules, like those shown in Table 1. It adds beliefs to memory that correspond to matched predicates. More abstract conceptual rules match against these inferences to produce higher-level beliefs. This bottom-up process continues until it computes the deductive closure of the agent's conceptual rules and its percepts.

Skill execution carries out complex activities over time. It operates in a top-down manner. On each cycle, Icarus finds the highest priority goal that is not satisfied by current beliefs. It selects a skill that includes this goal among its effects, and whose conditions match current beliefs. If the skill is primitive, the architecture carries out the associated actions, altering the environment. If the skill is nonprimitive, Icarus selects the first unachieved subgoal in its subskills clause, retrieves a skill with matched conditions that should achieve it, and either applies it (if primitive) or chains further (if not). Icarus executes only one step through the skill hierarchy on each cycle, waiting for new inferences before taking additional actions.

Icarus invokes problem solving when skill execution cannot find an applicable skill relevant to the current goal. The mechanism is a form of means-ends analysis (Newell, et al, 1958) that attempts to compose known skills and execute them as they become applicable. The process involves several capabilities: operator selection to pick skills that achieve a desired effect; operator subgoaling to decompose problems into simpler steps; and concept chaining to decompose goals into component conditions that can be individually achieved. The problem solver takes one of these steps on every cycle, continuing until it has satisfied all current goals, or it finds a skill ready for execution (Icarus interleaves problem solving with execution). The problem solver backtracks if it none of these steps applies.

Icarus also has the ability to perform one step look-ahead on the effects of skills, and to compute the deductive closure of these anticipated effects. The problem solver currently insists on greedy progress, ruling out skills that would undo achieved goals.

Results

We describe the results of applying Icarus to the sibling interaction problem by highlighting key elements of a problem solving trace.

The first step identifies the highest priority goal in the initial problem statement. Given that Sister1 begins in the comforted state, the only unachieved goal is:

(COMFORTED BROTHER1 ?toy1).

Since no skill lists the comforted relation in its effects, the system employs concept chaining to resolve this goal into its component parts. This leads to a new problem statement (after picking bindings, and removing true conditions):

((HAS-TOY BROTHER1 DOLL1))

Further concept chaining refines this statement into

((HOLDING BROTHER1 DOLL1)),

which finally makes contact with the effects of available skills. Icarus proposes to execute the skill:

(TAKE-DOLL BROTHER1 DOLL1).

as operator subgoaling reveals its conditions are all satisfied in the current state.

At this point, the projection component of the problem solver performs one-step look ahead with deductive closure on the effects of TAKE-DOLL, which are:

(HOLDING BROTHER1 DOLL1)
(NOT (HOLDING SISTER1 DOLL1))

This results in a hypothetical state where:

(COMFORTED BROTHER1 DOLL1)
(JEALOUS SISTER1 DOLL1)

are true, but

(COMFORTED SISTER1 DOLL1)

is not. In response, the problem solver rejects the TAKE-DOLL skill, as it undoes a goal present in the initial state. The problem solver backtracks at this point, eventually choosing the GET-DOLL skill, leading to a state where the brother and sister are both comforted:

(COMFORTED BROTHER1 BALL1)
(COMFORTED SISTER1 DOLL1)

Discussion

The sibling interaction scenario demonstrates several basic elements of social agency. It shows that the brother:

- models the sister agent's knowledge,
- considers interactions between their mental states,
- projects consequence of action on mental states, and
- generates action sequences that produce mental states.

At the same time, the approach has several obvious limitations. We touch on three of these, below, that relate to representation, reasoning, and planning and problem solving.

One obvious criticism is that the sibling interaction scenario bends a relational framework into service where modal constructs would be more convenient. For example, we would rather say that the brother believes the sister would be comforted in some circumstance than specialize two versions of the comforted relation. The end result is that the brother maintains a significant amount of knowledge about the sister in an inconvenient form.

A second criticism is that the system reasons about the sister's mental state by deduction from the brother's percepts, i.e., in an open loop without evidence obtained from the sister.

A third criticism is that planning and problem solving back-chains entirely from the brother's goals. For example, the sister's behavior to drop the doll on demand is only considered in service to the brother's objectives. The system needs to consider the sister's independent motivations instead, as well as her likely actions and reactions to the brother's choices.

In summary, the sibling interaction scenario illustrates that first steps towards social agency are feasible on the basis of current technology. It also illustrates that relational knowledge representations are limiting, that reasoning needs attention to issues of assessment and inference from observation, and that planning and problem solving needs a more thorough basis for envisioning future state. A research agenda on social agency needs to address these issues (among others) in general form.

A Proposed Research Agenda

While our initial results suggest that social agency is a feasible research goal, they leave many questions unanswered. We identify the key issues that future research in this area should address, and discuss promising approaches, below.

1. Develop representations that can express the complex relations among mental states required by social agency: These statements include modal assertions linking the beliefs, goals, intentions, and motivational states of multiple agents, e.g., "I believe you know the location of the ball", "I hold the goal that you are in the "comforted" state", and "I want Agent B to care if Agent C is jealous of agent D". We noted earlier that current modal logics lack sufficient power and convenience. One plausible approach is to specialize modal statements by encoding multiple levels (e.g., goals about beliefs about goals), and to organize those structures (e.g., into separate contexts or partitions). This will embed semantics in the structure (enhancing power) and improve convenience.

2. Develop methods of inferring user mental state from available observations. As mentioned earlier, this task fundamentally concerns abduction, which admits a variety of mechanisms, such as the use of evidential methods (e.g., Markov logics), or default logics to generate assumptions with belief revision to recover from faulty guesses. An alternate approach is to merge the kinds of case-based methods utilized in story understanding with the partitioned modal structures outlined above. Either path will require substantial new research to infer the complex conditions relating intent, preferences, and the affective states of multiple agents.

3. Predict the behavior of other agents given the situation and their mental state. This task admits two main approaches. One option is to predict the behavior of other agents by employing shallow models that tie a situation rather directly to action, through use of rules, cases, or predefined reactive strategies. Alternatively, we can develop nuanced models, for example, by employing all the facilities of a cognitive architecture to simulate an agent's reasoning. This context raises the interesting distinction between reasoning as that other agent using its inferred motivational structures (following a theory of mind), or utilizing its inferred knowledge in conjunction with own motivational state (loosely speaking, following simulation theory (Goldman, 2006)). We note that the simulation theory approach does not require inferring the motivational structure of other agents.

4. Employ knowledge about other agents to generate own agent behavior. Given the ability to predict other agent behavior, this issue involves two additional tasks; predicting the impact of own action on the mental state of other agents, and finding action that achieves desired mental state in other agents given their likely behavior. The first task implies a relatively modest research agenda to associate mental effects with actions, and to perform the reasoning necessary to ascertain implied effects. The second requires research to integrate problem solving with behavior prediction, and to control the degree of look ahead/search in a principled way.

5. Combine reasoning, planning and problem solving within an execution process. Cognitive architectures provide execution processes that combine their core cognitive functions, but social agency imposes additional demands for inference, assessment, and prediction. This implies a need for efficient control over the component activities. Relevant approaches include fixed interleaving policies (e.g., always act if possible, else problem-solve), and explicit meta-control.

Finally, we note that the issues listed above all concern the mechanics of behavior generation. The broader agenda is to employ those facilities to enable novel capability and application. For example, we can investigate computational models of altruism by examining the consequences of placing direct value on the mental states of other agents, or caring about them indirectly via the states they enable in one's self. We can conduct research on ethical models that associate value with actions, effects on self and others, or goals/motivations (loosely corresponding to deontological, consequential, and virtue ethics). The ability to interact over mental state presents an opportunity for work on collaborative problem solving (e.g., to partition tasks among participants according to their knowledge and skills). Lastly, as part of considering independently motivated user behavior, we have an opportunity to investigate emotional models, and their impact on action and as well as situation interpretation. Several of these investigations represent longer-term agenda items, but that research will be enabled by the pursuit of social agency.

We close by suggesting two challenge tasks that seem to require implementations of social agency. The first is to build a conversational weight-loss coach that motivates a non-motivated person to meet objectives. This task requires a degree of insight into the mental organization of the human user, as well as an element of sensitivity to his/her likely reactions. The second is to build an agent that runs a business meeting that must output a joint action plan. This task emphasizes intention setting, collaborative problem solving, and managing the agendas of multiple participants. Success at either challenge implies substantial progress towards social agency.

Concluding Remarks

This paper has presented a challenge problem for research on social agency, viewed as first step towards implementing broad social cognition. In the spirit of a challenge paper, we argued that this line of research is essential, significant, and feasible/timely. In particular, we claimed that social cognition is essential for addressing human level intelligence, that social agency motivates a novel line of research on achieving mental state with interesting application potential, and we showed that first steps toward social agency are feasible now by developing an initial demonstration in the context of the Icarus cognitive architecture.

Said broadly, cognitive systems with social agency will function more as social actors and collaborative problem solvers than as engineering tools. This, in turn, will call on us to reframe our model of how cognitive systems should be used.

References

Anderson, R., and Lebiere, C. (1998). *The atomic components of thought*. Lawrence Erlbaum Associates. ISBN 0-8058-2816-8

Byrne, R. W., Whiten, A. eds. (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Clarendon Press, Oxford.

Choi, D. *Reactive Goal Management in a Cognitive Architecture.* (2010). AAAI-10 Workshop on Goal-Directed Autonomy (GDA).

Dietterich, T. (1990). Exploratory research in machine learning. *Machine learning, 5, 5-10*.

Garvey, T., Kolojejchick, J., Angiolillo, C., Cummings, T., Gervasio, M., Haines, W., Knittel, J., Morley, D., Myers, K., Ommert, W., and Potter, S. (2011). *Learning by Demonstration Technology for Military Planning and Decision Making: A Deployment Story*. Proceedings of the Twenty-Third Conference on Innovative Applications of Artificial Intelligence.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford University Press. ISBN-10: 0195138929

Gratch, J. and Marsella, S. (2001). *Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents*. Proceedings of the 5th International Conference on Autonomous Agents, Montreal, Canada, June 2001

Greer, J., McCalla, G., eds. *Student Modeling: the Key to Individualized Knowledge --based Instruction*. NATO ASI Series, Vol. 125. Springer-Verlag, (1994) 167-189.

Horswill, I. (2008). *Lightweight procedural animation with believable physical interaction*. Proceedings of the Fourth Conference on Artificial Intelligence and Interactive Digital Entertainment. Stanford CA: AAAI Press.

Hsu, Feng-hsiung (2002). *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*. Princeton University Press. ISBN 0-691-09065-3

Klenk, M. (2010) Goal-Driven Autonomy in Planning and Acting. AAAI-10 Workshop on Goal-Directed Autonomy (GDA).

Langley, P., Choi, D., and Rogers, S. (2009). Acquisition of hierarchical reactive skills in a unified cognitive architecture. *Cognitive Systems Research, 10 , 316–332*.

Laird, J., Newell, A., and Rosenbloom, P. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33, 1–64*.

Leuski, A., and Traum, D. (2011). NPCEditor: Creating Virtual Human Dialogue Using Information Retrieval Techniques. *AI Magazine Volume 32 Number 2.*

Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65, 151–166*.

Rao, A., and Georgeff, M. (1995). *BDI Agents: From Theory to Practice*. Proceedings of the First International Conference on Multi-Agent Systems, ICMAS-95 San Francisco, USA

Shapiro, D., *Value-Driven Agents*. (2001). Ph.D. thesis, Stanford University, Department of Management Science and Engineering.

Swartout, W. (2010). Lessons Learned from Virtual Humans. *AI Magazine Volume 31 Number 1*.

Wilensky, R. (1983). *Planning and Understanding*. Addison Wesley Publishing. ISBN 0-201-09590-4.

Wooldridge, M. (2000). *Reasoning about Rational Agents*. MIT Press, Cambridge, Mass., 2000 ISBN 0–262–23213

Wray, R., Laird, J., Nuxoll, A., Stokes, D., and Kerfoot, A. *Synthetic adversaries for urban combat training*. Proceedings of the 2004 Innovative Applications of Artificial Intelligence Conference, San Jose, CA, July 2004.