Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)

Reasoning about Actions and Change: From Single Agent Actions to Multi-agent Actions (Extended Abstract)

Chitta Baral
Faculty of Computer Science and Engineering
Arizona State University
Tempe, AZ 85281, USA.

Abstract

We often deal with dynamic worlds where actions are executed by agents and events may happen. Example of such worlds range from virtual worlds such as the world of a database to robots and humans in physical worlds. To understand the dynamics of such worlds as well as to be able to assert some control over such worlds one needs to reason about the actions and events and how they may change the world. In this invited talk we will present some of the important results in this field and present some future directions. In particular, we will discuss how theories and results from reasoning about actions and change can be combined with theories and results in dynamic epistemic logics to obtain a unified theory of multi-agent actions.

Actions when executed often change the state of the world. Reasoning about actions helps us to predict if a sequence of actions is indeed going to achieve some goal that we may have; it allows us to plan or come up with a sequence of actions that would achieve a particular goal and maintain particular trajectories; it allows us to explain observations in terms of what actions may have taken place; and it allows us to diagnose faults in a system in terms of finding what actions may have taken place to result in the faults. When actions have non-deterministic effects, reasoning about actions is needed to verify policies and come up with policies to achieve goals and maintain desired trajectories. Thus, reasoning about actions is an important topic in Computer Science in general and in AI in particular. It has also served as a benchmark domain for evaluating knowledge representation languages.

Since the number of world states is often exponential in terms of the number of fluents (or individual properties of the world), a key aspect in reasoning about actions is to develop languages for succinct specification of the actions and their effects and define (as the semantics) the transition between “states” due to execution of actions. In a single agent world when the agent has complete observability “states” can be thought of as states of the world. In the presence of incompleteness of knowledge and sensing actions, “states” can be pairs consisting of states of the world and knowledge state of the agent.

There has been a lot of research in developing languages (often referred to as action description languages) that describe the world and the effect of actions, defining the above mentioned transitions with respect to action descriptions, and in using them for various kinds of reasoning tasks such as prediction, planning, counterfactuals, diagnosis, and proving correctness of plans, policies and execution programs. The importance of this was mentioned as early as 1969 by McCarthy (McCarthy and Hayes 1969). However, the systematic research approach of the last 20 years were guided and influenced by the high level action language approach (starting with $\mathcal{A}$ (Gelfond and Lifschitz 1992)), the approach of Sandewall (Sandewall 1994) and the approach at Toronto (Levesque et al. 1997; Reiter 2001). We made several contributions to this endeavor, such as (Baral and Gelfond 1993; 1997; Baral 1995; 1997; Son and Baral 2001; Baral, Tran, and Tuan 2002; Tran and Baral 2004).

In the first part of this invited talk we will briefly recall some of the key milestones in this journey to find ways to succinctly and naturally represent actions and reason about them.

The second part of the talk will be motivated by noticing that in most of these research[1] it was sufficient to assume that the actions were executed by a single agent; Even when multiple agents were referred to, the interactions between them were somewhat simple such as two agents simultaneously lifting a large table. However, in the real world much more involved interactions happen between multiple agents. A particularly hexing issue is when multiple agents are executing actions is the discrepancy of knowledge (and belief) between the agents and their knowledge (and belief) about each other’s knowledge (and belief). Although this issue has not been studied much in the reasoning about actions community, it has been studied in a somewhat different setting in dynamic epistemic logic (van Ditmarsch, van der Hoek, and Kooi 2007) and logics for epistemic programs (Baltag and Moss 2004) communities.

Inspired by those work, we will start the second part of the talk with a simple multi-agent scenario where an action execution involves three classes of agents with respect to what

---

[1] There were a few exceptions such as (Ghaderi, Levesque, and Lesprance 2007; Gelfond 2007).

they know about this execution: (a) the group of agents that are fully aware of the execution of the action; (b) the group of agents that have some clue about the execution of the action, but not the complete information; and (c) the group of agents that have no clue about the execution of the action.

Our first example is inspired by examples in (Baltag and Moss 2004). Consider a box to which a coin was tossed in presence of agents a, b and c but none were close enough to see how the coin fell and immediately the box cover got closed. Thus it becomes common knowledge between the agents a, b, and c that none of them know the value (of the top side) of the coin. This state of affairs can be expressed by Figure 1 which actually represents two "states": (s_1, M) and (s_2, M), where M is the Kripke model shown in the figure and s_1 denotes $\{H\}$ and s_2 denotes $\{T\}$.

Now consider the action of a peeking at the box and finding out that the top of the coin is head, b seeing from far that a peeked but not knowing what a saw, c being distracted and having no idea about what happened, and a and b having the common knowledge about all of this.

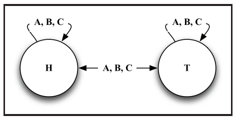

Figure 1: A Kripke structure describing the beliefs of agents in the "Coin World."

A key question is: What changes does the above action cause? In other words, how do we define the transition caused by the above action with respect to the "states" (s_1, M) and (s_2, M).

The following figure shows the transition.

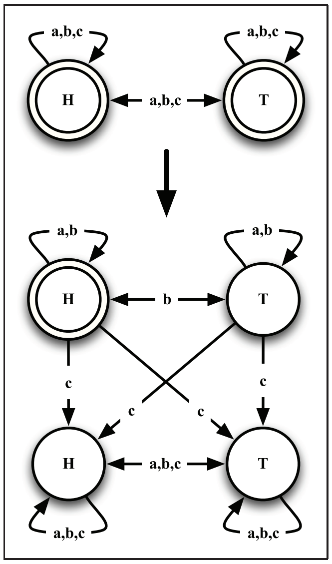

Figure 2: Transition due to the action a knowing H, b noticing from far and c having no clue

The transition shown in Figure 2 allows the following expected conclusions.

- Before the action: Agents a, b and c do not know H to be true nor do they know T to be true.
  After the action: Agents b and c do not know H to be true nor do they know T to be true. But agent a knows that H is true.

  The above captures the aspect of the action that a peeks and finds out that H is true.
- Before the action: Agents a and c know that agent b does not know H to be true.
  After the action: Agents a knows that agent b does not know H to be true; agent c believes that agent b does not believe H to be true.

  The above captures the aspect of the action that b has not peeked and a knows that and c believes that.
- Before the action: Agents b and c know that agent a does not know H to be true.
  After the action: Agent b knows that a may know H to be true. Agent c incorrectly believes that agent a does not know H to be true.

  The above captures the aspect of the action that b sees (from far) a peeking and c having no clue.
- Before the action: Agent a knows that agent b knows that agent a does not know H to be true.
  After the action: Agent a knows that agent b knows that agent a may know H to be true.

  The above captures the aspect of the action that a is aware that b is observing him peeking.
- Before the action: Agent a knows that agent c knows that agent a does not know H to be true.
  After the action: Agent a knows that agent c believes that agent a does not know H to be true.

  The above captures the aspect of the action that a is aware that c is clueless about a's peeking.

We believe that defining such transitions with respect to a basic set of multi-agent actions would form the foundation of multi-agent reasoning about actions. Following are some examples of basic actions involving multiple agents that change the agents' knowledge about the world as well as about each others' knowledge.

- Agent a is told the value of fluent f, agent b observes that a is told the value of fluent f but does not know what exactly was told and agent c has no clue.
- Agent a senses the value of fluent f, agent b observes that a is sensing the value of fluent f and agent c has no clue.
- Agent a may have sensed the value of fluent f, agent b notices that a may have sensed the value of fluent f and agent c has no clue.
- Agent a has no idea about f or $\neg f$ but lies to b that it is f and agent c has no clue about this lying.
- Agent a has no idea about f or $\neg f$ but misleads b that it knows and agent c has no clue about this misleading.
- Agent a has no idea about f or $\neg f$ but misleads b that it may know and agent c has no clue about this misleading.

Although such three level actions have not been considered in the earlier work on knowledge systems (Fagin et al. 1995), dynamic epistemic logic (van Ditmarsch, van der Hoek, and Kooi 2007) and logics for epistemic programs (Baltag and Moss 2004), these works provide some of the machinery that can be used to define the transition due to such actions. On the other hand research in reasoning about actions provide ways to succinctly represent worlds and transitions and account for causal connection between fluents. Thus putting together ideas and machineries from each of them would lead to a valuable theory of multi-agent actions and change. We took an initial step in that direction in (Baral et al. 2010).

Representing and reasoning about such multi-agent actions is not an academic exercise. For example, they lead to some interesting planning applications, especially in a battle field scenario. In a battle field an agent can often classify the other agents as close friends, far-off friends and foes. The first kind of friend correspond to agent a of the above mentioned example and they exactly know what happened, while the second kind correspond to agent b who are partly aware of what happened but for their own good are not informed of everything. The foe correspond to agent c who are perhaps forced to be distracted so that they have no clue of what happened. Based on the above, the following is a planning scenario where reasoning about other agents's knowledge plays an important role.

A planning scenario: The agent a has been imprisoned by agent c in a cell and c has hidden the key. Agent a would like to escape from the prison but needs b's help in that. Agent c is patrolling the area, but there are small windows of time when b can come in and help. But the time windows are so small that b can only help if it knows where the key is. With c patrolling it will not have enough time to come and look for the key. In such a scenario a makes the following plan. It will first find out where the key is without c knowing but b being signalled that a knows where the key is. Now b can move in at the opportune moment, find from a where the key is and rescue him.

Towards the end of our talk we will discuss the role of answer set programming (ASP) (Gelfond and Lifschitz 1988; Baral 2003) and other knowledge representation languages in representing and reasoning about action and change. We will show that ASP can also be used to reason about about the knowledge and belief modalities that manifest in a multi-agent setting. We will conclude by discussing some additional future research directions.

References

Baltag, A., and Moss, L. 2004. Logics for epistemic programs. *Synthese*.

Baral, C., and Gelfond, M. 1993. Representing concurrent actions in extended logic programming. In *Proc. of 13th International Joint Conference on Artificial Intelligence, Chambery, France*, 866–871.

Baral, C., and Gelfond, M. 1997. Reasoning about effects of concurrent actions. *Journal of Logic Programming* 31(1-3):85–117.

Baral, C.; Gelfond, G.; Son, T.; and Pontelli, E. 2010. Using answer set programming to model multi-agent scenarios involving agents' knowledge about other's knowledge. AAMAS 2010.

Baral, C.; Tran, N.; and Tuan, L. 2002. Reasoning about actions in a probabilistic setting. In *Proc. of AAAI'2002*, 507–512.

Baral, C. 1995. Reasoning about Actions : Non-deterministic effects, Constraints and Qualification. In *Proc. of IJCAI 95*, 2017–2023.

Baral, C. 1997. Embedding revision programs in logic programming situation calculus. *Journal of Logic Programming* 30(1):83–97.

Baral, C. 2003. *Knowledge representation, reasoning and declarative problem solving*. Cambridge University Press.

Fagin, R.; Halpern, J.; Moses, Y.; and Vardi, M. 1995. *Reasoning about Knowledge*. MIT press.

Gelfond, M., and Lifschitz, V. 1988. The stable model semantics for logic programming. In Kowalski, R., and Bowen, K., eds., *Logic Programming: Proc. of the Fifth Int'l Conf. and Symp.*, 1070–1080. MIT Press.

Gelfond, M., and Lifschitz, V. 1992. Representing actions in extended logic programs. In Apt, K., ed., *Joint International Conference and Symposium on Logic Programming.*, 559–573. MIT Press.

Gelfond, G. 2007. A declarative framework for modeling multi-agent systems. Masters thesis, Texas Tech University.

Ghaderi, H.; Levesque, H.; and Lesprance, Y. 2007. A logical theory of coordination and joint ability. In *AAAI 2007*, 421–426.

Levesque, H.; Reiter, R.; Lesperance, Y.; Lin, F.; and Scherl, R. 1997. GOLOG: A logic programming language for dynamic domains. *Journal of Logic Programming* 31(1-3):59–84.

McCarthy, J., and Hayes, P. 1969. Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B., and Michie, D., eds., *Machine Intelligence*, volume 4. Edinburgh University Press. 463–502.

Reiter, R. 2001. *Knowledge in action: logical foundation for describing and implementing dynamical systems*. MIT press.

Sandewall, E. 1994. The range of applicability of some non-monotonic logics for strict inertia. *Journal of Logic and Computation* 4(5):581–616.

Son, T., and Baral, C. 2001. Formalizing sensing actions: a transition function based approach. *Artificial Intelligence* 125(1-2):19–93.

Tran, N., and Baral, C. 2004. Encoding probabilistic causal models in probabilistic action language pal. In *AAAI'04*.

van Ditmarsch, H.; van der Hoek, W.; and Kooi, B. 2007. *Dynamic Epistemic Logic*. Springer.