From Human-Human Joint Action to Human-Robot Joint Action

Contents

1.1 Joint Action Theory

A first step to endow robots with the ability to perform Joint Actions with humans is to understand how humans act together. As a working definition of Joint Action, we will use the one from [Sebanz 2006]:

> *Joint action can be regarded as any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment.*

A given number of prerequisites are needed for these individuals to achieve the so-called Joint Action. First of all, they need to agree on the change they want to bring in the environment, the conditions under which they will stay engaged in its realisation and the way to do it. A number of works have studied this prerequisite, named *commitment*, which I will develop in Sec. 1.1.1. Then, as mentioned in the

definition, the individuals need to coordinate their actions in space and time. This will be studied in Sec. 1.1.3. Finally, in order to coordinate, each individual needs to be aware of the other, he needs to be able to perceive him and predict his actions. This part will be develop in Sec. 1.1.2.

1.1.1 Commitment

The first prerequisite to achieve a Joint Action is to have a *goal* to pursue and the *intention* to achieve it. Let's define in a first time what is called a *goal* and an *intention* for a single person before going to a *joint goal* and a *joint intention*.

In [Tomasello 2005], Tomasello et al. define what they call a *goal* and an *intention* and illustrate these definitions with an example and an associated figure (fig. 1.1) where a person wants to open a box.
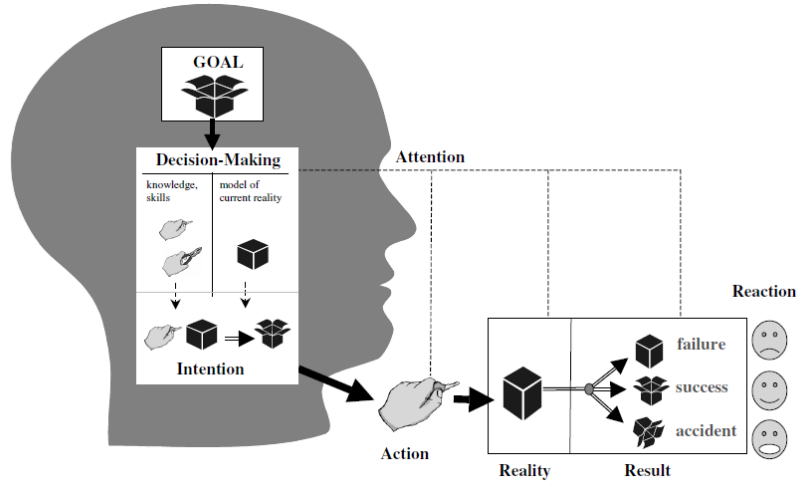

Figure 1.1: Illustrative example of an intentional action by Tomasello et al. Here the human has for *goal* to open the box. He chooses a means to perform it and so forms an *intention*.

A *goal* is defined here as the representation of the desired state by the agent (in the example, the goal is an open box) and, based on Bratman's work [Bratman 1989], an *intention* is defined as an action plan the agent commits itself in pursuit of a goal (in the example, the intention is to use a key to open the box). The *intention* includes both a *goal* and the means to achieve it.

In a same way, Cohen and Levesque propose in [Cohen 1991] a formal definition of what they call a *persistent goal*:

> **Definition:** An agent has a *persistent goal* relative to q to achieve p iff:
>
> 1. she believes that p is currently false;
> 2. she wants p to be true eventually;

3. it is true (and she knows it) that (2) will continue to hold until she comes to believe either that p is true, or that it will neither be true, or that q is false.

However, their definition of an *intention* differs a little from the previous one. They define an *intention* as a commitment to act in a certain mental state:

Definition: An agent *intends* relative to some conditions to do an action just in case she has a persistent goal (relative to that condition) of having done the action, and, moreover, having done it, believing throughout that she is doing it.

The *intention* still includes the *goal* but here it concerns more the fact that the agent commits itself to achieve the goal than the way to achieve it.

Let's now apply these principles to a Joint Action. One of the most known definition of *joint intention* is the one of Bratman [Bratman 1993]:

We intend to J if and only if:

1. (a) I intend that we J and (b) you intend that we J.

2. I intend that we J in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b; you intend that we J in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b.

3. 1 and 2 are common knowledge between us.

This definition is taking back and illustrated by Tomasello et al. in [Tomasello 2005] where they reuse the example of the box to open (fig 1.2).

The *shared goal* is defined as the representation of a desired state plus the fact that it will be done in collaboration with other person(s) (in the example, they will open the box together) and a *joint intention* is defined as a collaborative plan the agents commit to in order to achieve the *shared goal* and which takes into account both agents individual plans (here an agent will hold the box with the clamp while the other open it with the cutter).

In a same way, Cohen and Levesque extend their definition of *persistent goal* and *intention* to a collaborative activity. They first define a *weak achievement goal* as:

Definition: An agent has a *weak achievement goal* relative to q and with respect to a team to bring about p if either of these conditions holds:

- The agent has a normal achievement goal to bring about p, that is, the agent does not yet believe that p is true and has p eventually being true as goal.
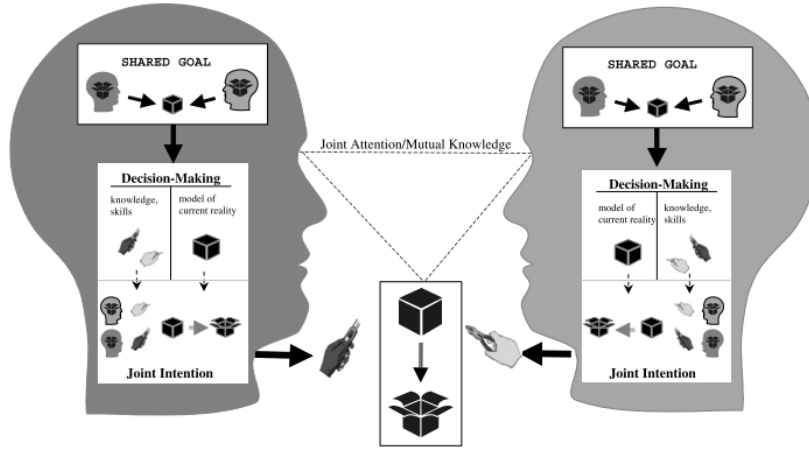
Figure 1.2: Illustrative example of a collaborative activity by Tomasello et al. Here the humans have for *shared goal* to open the box together. They choose a means to perform it which takes into account the other capabilities and so form a *joint intention*.

- The agent believes that p is true, will never be true, or is irrelevant (that is, q is false), *but* has as a goal that the status of p be mutually believed by all the team members.

They then use this definition to define a *joint persistent goal*:

Definition: A team of agents have a *joint persistent goal* relative to q to achieve p just in case

- they mutually believe that p is currently false;
- they mutually know they all want p to eventually be true;
- it is true (and mutual knowledge) that until they come to mutually believe that p *is true*, that p will never be true, or that q is false, they will continue to mutually believe that they each have p as a weak achievement goal relative to q and with respect to the team.

They finally define a *joint intention* as:

Definition: A team of agents *jointly intends*, relative to some escape condition, to do an action iff the members have a joint persistent goal relative of that condition of their having done the action and, moreover, having done it mutually believing throughout that they were doing it.

As previously, the definitions of Cohen and Levesque do no take into account the way to achieve the *shared goal*, however, they introduce the interesting idea that agents are also engaged to communicate about the state of the *shared goal*.

Concerning the way to achieve a *shared goal*, mentioned into the definition of the *joint intention* of Tomasello et al., Grosz and Sidner initially introduce and formalize

the notion of *Shared Plan* in [Grosz 1988], which is extended in [Grosz 1999]. The key properties of their model are as follows:

1. it uses individual intentions to establish commitment of collaborators to their joint activity

2. it establishes an agent's commitments to its collaborating partners' abilities to carry out their individual actions that contribute to the joint activity

3. it accounts for helpful behavior in the context of collaborative activity

4. it covers contracting actions and distinguishes contracting from collaboration

5. the need for agents to communicate is derivative, not stipulated, and follows from the general commitment to the group activity

6. the meshing of subplans is ensured it is also derivative from more general constraints.

With their definition, each agent does not necessarily know the all *Shared Plan* but only his own individual plan and the meshing subparts of the plan. The group has a *Shared Plan*, but no individual member necessarily has the whole *Shared Plan*.

In conclusion, the concepts concerning the commitment of agents to a collaborative activity that we will use in this thesis can be summarized as:

- A *goal* will be represented as a desired state.

- A *shared goal* will be considered as a *goal* to be achieved in collaboration with other partner(s). An agent is considered engaged in a *shared goal* if he believes the goal is currently false, he wants the goal to be true and he will not give up on the goal unless he knows that the goal is achieved, not feasible or not relevant any more and he knows that his partners are aware of it.

- A *joint intention* will include a *shared goal* and the way to realize it. This way will be represented as a *Shared Plan* which will take into account each agent capacities and the potential conflicts between their actions. This *Shared Plan* will not be necessarily completely known by all members of the group but all individuals will know their part of the plan and the meshing subparts.

1.1.2 Perception and prediction

One important thing for an agent when performing a Joint Action is to be able to perceive and predict the actions of his partner and their effects. Based on the works in [Sebanz 2006], [Pacherie 2011] and [Obhi 2011] we identified several necessary abilities for this predictions:

- **Joint attention:** The capacity for an agent to direct his attention to where the one of his partner is directed allows to share a representation of objects and events. It brings a better understanding of the other agent knowledge and where his attention is focused and so, it helps the prediction of his possible next actions. Moreover, there should be a mutual manifestation of this joint attention, meaning that we should show that we share the other attention.

- **Action observation:** Several studies have shown that when someone observes another person executing an action, a corresponding representation of the action is formed for the observer [Rizzolatti 2004]. This is done by what has been called the *mirror-neuron* system. This behavior allows the observer to predict the outcomes of the actor's action.

- **Co-representation:** An agent needs to have a representation of his partner, including his goal, his capacities and the social rules he is following. Indeed, having this representation will help to predict his future actions. For example, a pedestrian who sees a red traffic light will be able to predict that the car drivers will stop. In a same way, if you know that someone is aiming to go out shopping, you will be able to predict that he will look for the key of the car.

- **Agency:** Sometimes, when there is a close link between an action performed by oneself and an action performed by another, it can be hard to distinguish who caused a particular action effect. The capacity to attribute the action effects to the good actor is called the sense of *Agency*. This sense of *Agency* is an important thing in Joint Action in order to correctly predict the effects of each action.

Based on the same works as before and on [Sebanz 2009], we can list several kinds of predictions to support Joint Action which can be done thanks to the abilities described previously :

- **What:** A first one is to predict what will do an agent. Two kinds of predictions can be distinguished here:

  - *action-to-goal:* this is supported by the *mirror-neuron* system introduced before. Here the therm goal concerns the goal of an action, its purpose. The idea is that observing an action, it is possible to predict its goal. For example, if we observe someone extending his arm toward an object we can predict that he will pick the object.

  - *goal-to-action:* here the therm goal concerns the goal of a task, as defined in the previous subsection. Knowing this goal, it can be easy to predict which action an agent will perform.

- **When:** another prediction which is necessary is the timing of an action. Knowing when an action will occur and during how long allows to a better coordination in time.

- **Where:** a Joint Action usually takes place in a shared space. It is therefore necessary to predict the future position of the partner and his actions in order to coordinate in space.

1.1.3 Coordination

Knoblich [Knoblich 2011]

- emergent coordination: entertainment, affordances [Gibson 2014], corrspondance action-perception, action simulation

- planned coordination: coordination smoother (vesper, [Vesper 2010])

+communication: Clark [Clark 1996]

1.2 How to endow a robot with Joint Action abilities

1.2.1 Engagement and Intention

- Intention and plan recognition

- Goal reasoning

- Engagement in the task

- human-aware task planning

1.2.2 Perspective taking and humans mental states

- world state management

- perspective taking

- mental states

1.2.3 Actions realization

- human-aware motion planning

- human-aware control

- understanding effects of actions (agency)

1.2.4 Coordination

- plan execution

- dialogue

- signalling

- joint actions (e.g. handover)

1.3   A three levels architecture

1.3.1   The three levels of Pacherie

[Pacherie 2008], [Pacherie 2011]

1.3.2   A three levels robotics architecture

+ related work on others robotics architecture

Bibliography

[Bratman 1989] Michael E Bratman. *Intention and personal policies*. Philosophical perspectives, vol. 3, pages 443–469, 1989. (Cited in page 2.)

[Bratman 1993] Michael E Bratman. *Shared intention*. Ethics, vol. 104, no. 1, pages 97–113, 1993. (Cited in page 3.)

[Clark 1996] Herbert H Clark. *Using language. 1996*. Cambridge University Press: Cambridge, vol. 952, pages 274–296, 1996. (Cited in page 7.)

[Cohen 1991] Philip R Cohen and Hector J Levesque. *Teamwork*. Nous, vol. 25, no. 4, pages 487–512, 1991. (Cited in page 2.)

[Gibson 2014] James J Gibson. *The theory of affordances*. Jen Jack Gieseking, William Mangold, Cindi Katz, Setha Low, and Susan Saegert, editors, The People, Place, and Space Reader, pages 56–60, 2014. (Cited in page 7.)

[Grosz 1988] Barbara J Grosz and Candace L Sidner. *Plans for discourse*. Technical Report, DTIC Document, 1988. (Cited in page 5.)

[Grosz 1999] Barbara J Grosz and Sarit Kraus. *The evolution of SharedPlans*. In Foundations of rational agency, pages 227–262. Springer, 1999. (Cited in page 5.)

[Knoblich 2011] Günther Knoblich, Stephen Butterfill and Natalie Sebanz. *3 Psychological research on joint action: theory and data*. Psychology of Learning and Motivation-Advances in Research and Theory, vol. 54, page 59, 2011. (Cited in page 7.)

[Obhi 2011] Sukhvinder S Obhi and Natalie Sebanz. *Moving together: toward understanding the mechanisms of joint action*, 2011. (Cited in page 5.)

[Pacherie 2008] Elisabeth Pacherie. *The phenomenology of action: A conceptual framework*. Cognition, vol. 107, no. 1, pages 179–217, 2008. (Cited in page 8.)

[Pacherie 2011] Elisabeth Pacherie. *The Phenomenology of Joint Action: Self-Agency versus Joint Agency*. Joint attention: New developments in psychology, philosophy of mind, and social neuroscience, page 343, 2011. (Cited in pages 5 and 8.)

[Rizzolatti 2004] Giacomo Rizzolatti and Laila Craighero. *The mirror-neuron system*. Annu. Rev. Neurosci., vol. 27, pages 169–192, 2004. (Cited in page 6.)

[Sebanz 2006] Natalie Sebanz, Harold Bekkering and Günther Knoblich. *Joint action: bodies and minds moving together*. Trends in cognitive sciences, vol. 10, no. 2, pages 70–76, 2006. (Cited in pages 1 and 5.)

[Sebanz 2009] Natalie Sebanz and Guenther Knoblich. *Prediction in joint action: What, when, and where.* Topics in Cognitive Science, vol. 1, no. 2, pages 353–367, 2009. (Cited in page 6.)

[Tomasello 2005] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne and Henrike Moll. *Understanding and sharing intentions: The origins of cultural cognition.* Behavioral and brain sciences, vol. 28, no. 05, pages 675–691, 2005. (Cited in pages 2 and 3.)

[Vesper 2010] Cordula Vesper, Stephen Butterfill, Günther Knoblich and Natalie Sebanz. *A minimal architecture for joint action.* Neural Networks, vol. 23, no. 8, pages 998–1003, 2010. (Cited in page 7.)