Collective Intention Revision from a Database Perspective

No Author Given

No Institute Given

Logics of rational agency try to capture, in logic, various facets of human mental state, and place normative relationships among them. One category of mental state includes *informational attitudes*, and notions such as beliefs and knowledge fall into this category. In the past thirty years, logics of knowledge and beliefs have become an industry, ranging from philosophy, to computer science and economics (cf. [2, 6, 8]). Other mental states are *motivational attitudes* that capture something about the agents' preference structure, and *action attitudes* that capture his inclination towards taking certain actions. In a typical theory, the action attitudes mediate between the informational and motivational attitudes; the agent's choice of action is dictated by his wants and beliefs.

When studying these different attitudes in isolation, things are relatively easy. But this becomes more involved when one considers the interaction between the various types of attitude. For example, the dynamics of belief and preference are in general intertwined, with changes in beliefs leading to change in preference (and possibly vice versa). But more complex interactions are common. Suppose for instance that someone has the *goal* to attend a conference, which was given rise to by the *desire* to publish. One then may *intend* to take certain actions in service of this goal, such as writing a paper, submitting a paper, booking a flight, reserving a hotel. These intentions are motivated by the *belief* that the paper will be accepted at the conference, and they are inconsistent with believing that there is no budget to travel. Each subsequent intention leads to an increased level of *commitment* towards the goal to attend the conference.

This is a complicated picture, and so it's not surprising that progress on this has been relatively slow. In order to contribute to this, Shoham [10] recently proposed to take an *artifactual perspective* on the notion of intention. When taking this perspective, the focus is on a particular artifact (usually defined abstractly in mathematical terms), whose behavior is completely specified and thus in principle understood, but for which one seeks an intuitive high-level language to describe its behavior. A good example is the use of the notion of "knowledge" to reason about distributed systems [6]. The protocol governing the distributed system is well specified, but intuitively one tends to speak about what one processor does or does not know about another processor, and the role of the mathematical theory of knowledge is to formalize this reasoning. Shoham uses the artifactual perspective on intentions by choosing as an artifact a database, and he coins it the *database perspective*.

A database represents information in a specific format, and provides various services associated with this information, the most basic services being storage and retrieval. Logic can provide the epistemological theory of the database, capturing the semantics of the information stored in the database and of the operations on it. The most well-known theory for a belief database is probably the Alchourrón-Gärdenfors-Makinson (AGM) theory [1], which has been extremely influential in the area of belief

change in both computer science and philosophy in the past few decades. Such an "intelligent" AGM belief database captures the beliefs of the agent, and, in addition to the basic storage and retrieval operations, it ensures that the beliefs remain consistent at all times while requiring a notion of *minimal change*.[1]

In order to model an intention database in the same spirit, consider the intention database as being in service of some planner, in particular of the sort encountered in so-called "classical" AI planning [15]. The planner posits a set of actions to be taken at various times in the future, and updates this set as it continues with its deliberations and as it learns new facts. In the philosophical parlance, these are "future-directed intentions" (Figure 1). This framework is conceptually proposed by Shoham and subsequently formalized by Icard *et al.* [7], who introduce a notion of *coherence* between the belief database and the intention database: The belief database of an agent coheres with the intention database when the agent considers it possible to carry out all of the intentions. They then define AGM-like postulates for coherent belief and intention revision.

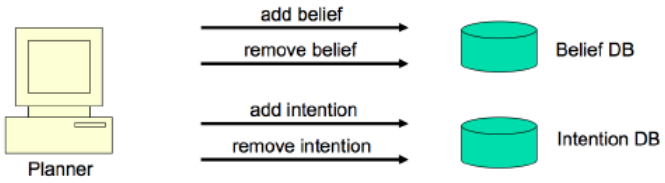

Fig. 1. The database perspective

We can model our conference scenario using the database perspective by supposing that the agent is planning to attend the conference (note that *goals* are explicitly modeled in the database perspective, but they are assumed to be part of the planner). In the course of planning, the agent will add intentions (which are understood as time-labeled actions) to his intention database, such as "write a paper", "submit the paper", and, at some point, "attend the conference". These intentions will have to cohere with his beliefs: It would be irrational for the agent to have the intention to visit the conference while he believes there is no traveling budget. Therefore, if the paper has been accepted and the agent finds out that there is no traveling budget, he will have to drop his intention to visit the conference.

The focus of the current paper is to study the database perspective in a *multi-agent setting*. Thus, returning to the example, suppose now there is a group researchers, each with a belief database and an intention database, who are planning to go to conferences, possibly by writing papers together. They have to coordinate their intentions in some way, while they might not know all of the others' beliefs and intentions. Moreover,

[1] We are fully aware of the shortcomings of the AGM framework, in particular when it comes to iterated revision. For now, we decide to stick to it for ease of exposition, but note that we plan future work containing more sophisticated means of modeling belief dynamics.

there may be dependencies between intentions of different researchers when they are collaborating on a paper.

In order to tackle this problem, we propose to extend the database perspective with a single auxiliary database in the multi-agent system that we coin the "Collective Intention Database". The importance of collective intentions has been recognized for some time (see, e.g. [4, 5, 11, 12] and many others). There is still much disagreement in the philosophical literature about whether collective intentions can be reduced to more basic intentions of the members of the group (cf. [3, 12, 13]), or whether collective intentions are in some way irreducible (cf. [9, 14]). Once again, appealing to the database perspective allows us to sidestep these debates and center our study on what promises to be most useful.

The collective intention database contains a set of collective intentions. A collective intention is formalized as a set of action-agent pairs and a time point. In this way, the collective intention database captures two basic notions. Firstly, it serves as a coordination mechanism for a set of agents by specifying what actions of the agents are carried out simultaneously, and in this way represents dependencies between the intentions of the individual agents. Secondly, it allows agents to reason about each others intentions, in the sense that each agent believes that the other agents will carry out their part in a collective intention. We extend the Icard *et al.* definition of coherence in a natural way: An agent's beliefs cohere with his intentions and the collective intentions if 1) the agent considers it possible to carry out all of his intended actions (this is equivalent to the Icard *et al.* definition), and 2) the agent considers it possible that all intentions by other agents in collective intentions in which the agent itself also participates are carried out by the corresponding agents. We use this notion of coherence to define AGM-like revision postulates for beliefs, individual intentions, and collective intentions. Our main result is a representation theorem, which states that we can equivalently represent the revision postulates for the belief, individual intention, and collective intention databases as revision operators on a semantic model for the multi-agent system.

References

1. Alchourron, C.E., Gardenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic 50(2), 510–530 (06 1985)
2. van Benthem, J.: Exploring Logical Dynamics. Center for the Study of Language and Information (1996)
3. Bratman, M.: Faces of Intention: Selected Essays on Intention and Agency. Cambridge Studies in Philosophy, Cambridge University Press (1999)
4. Bratman, M.E.: Shared intention. Ethics 104(1), 97–113 (1993)
5. Cohen, P.R., Levesque, H.J.: Teamwork. Noûs 25(4), 487–512 (1991)
6. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning about Knowledge. MIT Press (1995)
7. Icard, T., Pacuit, E., Shoham, Y.: Joint revision of belief and intention. Proc. of the 12th International Conference on Knowledge Representation pp. 572–574 (2010)
8. Peppas, P.: Handbook of Knowledge Representation, chap. Belief Revision. Elsevier (2007)
9. Searle, J.: The Construction of Social Reality. Free Press (1995)
10. Shoham, Y.: Logical theories of intention and the database perspective. Journal of Philosophical Logic (2009)

11. Subramanian, R.A., Kumar, S., Cohen, P.R.: Integrating joint intention theory, belief reasoning, and communicative action for generating team-oriented dialogue. In: AAAI. pp. 1501–1507. AAAI Press (2006)
12. Tuomela, R.: Joint intention, we-mode and i-mode. Midwest Studies In Philosophy 30(1), 35–58 (2006)
13. Tuomela, R., Miller, K.: We-intentions. Philosophical Studies 53(3), 367–389 (1988)
14. Velleman, J.D.: How to share an intention. Philosophy and Phenomenological Research 57-1, 29–50 (1997)
15. Weld, D.S.: Recent advances in ai planning. AI Magazine 20(2), 93–123 (1999)