Second-Order Know-How Strategies

Jia Tao
Department of Computer Science
Lafayette College
Easton, Pennsylvania 18042
taoj@lafayette.edu

In this paper[1] we study the interplay between coalition strategies and the distributed knowledge in multiagent systems.

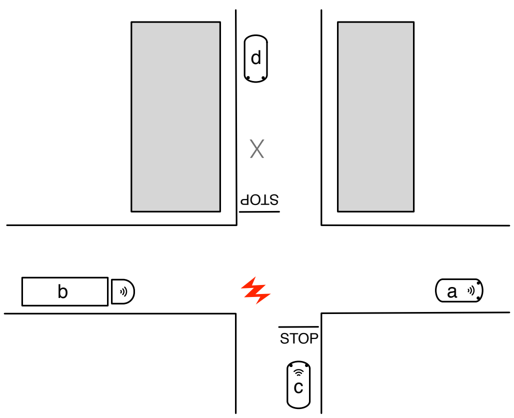

Figure 1: State of Traffic w_1.

Consider the traffic situation depicted in Figure 1, where a regular vehicle d and three self-driving vehicles a, b, and c approach an intersection. There are stop signs at the intersection facing cars c and d. According to the traffic rules, these two cars must slow down, stop, and yield to truck b.

Suppose that the driver of car d does not notice the sign and, as a result, this car is approaching the intersection with a constant speed. If neither of the vehicles changes its behavior, car d will hit the side of the rear half of truck b, at the location marked with a red zigzag shape on the figure. Truck b can potentially slow down, but then car d will hit the side of truck b in the front half instead of the rear half. Thus, to avoid being struck by car d, truck b has to accelerate and pass the intersection before car d does. We assume that in this case if car a maintains the same speed, then there will be a rear-end collision between truck b and car a. Hence, to avoid any collision, not only must truck b accelerate, but car a must accelerate as well. In other words, to prevent a collision, vehicles a and b must engage in a strategic cooperation. We say that coalition $\{a, b\}$ has a *strategy* to prevent a collision.

[1] This is a joint work with Pavel Naumov presented at 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 18), July 10-15, 2018, Stockholm, Sweden.

The traffic situation is further complicated by two buildings, shown in Figure 1 as grey rectangles. The buildings prevent car a and truck b from seeing car d. Although coalition $\{a, b\}$ has a strategy to avoid collision, it does not know what this strategy is, nor does it know that such a strategy exists. However, self-driving car c can observe that car d is not slowing down, and it can make coalition $\{a, b\}$ aware of the presence of car d as well as its speed and location. With the information shared by car c, not only will coalition $\{a, b\}$ have a strategy to avoid a collision, but it also will know what this strategy is.

In general, the following cases might take place: (i) a coalition does not have a strategy; (ii) a coalition has a strategy, but it does not know that it has a strategy; (iii) a coalition knows that it has a strategy, but it does not know what the strategy is; (iv) a coalition knows that it has a strategy and it knows what this strategy is. In the last case, we say that the coalition has a *know-how* strategy.

In our example, coalition $\{a, b\}$ has a know-how strategy to avoid a collision after car c shares the traffic information with the coalition. In other words, it is not coalition $\{a, b\}$ but the single-element coalition $\{c\}$ that knows what is the strategy of coalition $\{a, b\}$. We refer to such strategies as *second-order* know-how strategies by analogy with the commonly used term *second-order knowledge*.

Second-order know-how manifests itself in many settings. For example, a teacher might know how a student can succeed or a group of campaign advisers might know how a political party can win the elections.

We use the notion of an *epistemic transition system* to formalize the concept of a second-order know-how strategy. A fragment of an epistemic transition system corresponding to the traffic situation described above is depicted on the diagram in Figure 2. In particular, the traffic situation depicted in Figure 1 is represented by state w_1 on this diagram. The arrows on the diagram correspond to possible transitions of the system.

For the sake of simplicity, we assume that transitions of this system only depend on the actions of agents a and b. Moreover, each of agents a and b is assumed to have just three strategies: to slow down ($-$), to maintain current speed (0), and to accelerate ($+$). In Figure 2, transitions are labeled by the strategies of agents a and b that accomplish the transition.

If a transition is labeled with strategy profile (x, y), then x represents the strategy of car a and y represents the strategy of truck b. Although there are nine possible transitions from state w_1, corresponding to nine possible strategy profiles (x, y), the fragment of this system (depicted in Figure 2) shows only four such transitions leading to states w_4, w_5, w_6, and w_7.

As we discussed earlier, vehicles a and b can use coalition strategy $(+,+)$ to avoid a collision. However, they *do not know* that this coalition strategy would prevent a collision because they do not even know the presence of vehicle d, let alone its location and speed. To show this formally, consider a hypothetical state w_2 in Figure 2. In this state, vehicle d is currently at the spot marked by symbol X on Figure 1. In this state, car d is closer to the intersection than it is in state w_1. A simultaneous acceleration of vehicles a and b (coalition strategy $(+,+)$) would not prevent a collision because car d would hit the side of truck b in the rear half. Instead, in state w_2, coalition $\{a, b\}$ can use, for example, strategy $(0,-)$ to avoid a collision. Under this strategy, car a maintains the current speed and truck b slows down. Since vehicles a and b cannot see vehicle d, they cannot distinguish states w_1 and w_2. We define a coalition *know-how strategy* at state w_1 as a strategy that would succeed in all states indistinguishable from state w_1 by the coalition. Thus, the transition system whose fragment is depicted on Figure 2 does not have a know-how strategy for coalition $\{a, b\}$ to avoid a collision in state w_1. States w_1 and w_2 are not the only indistinguishable states in this system. For example, state w_3 in the same figure, where car d is not present at the scene, is also indistinguishable from states w_1 and w_2 to coalition $\{a, b\}$.

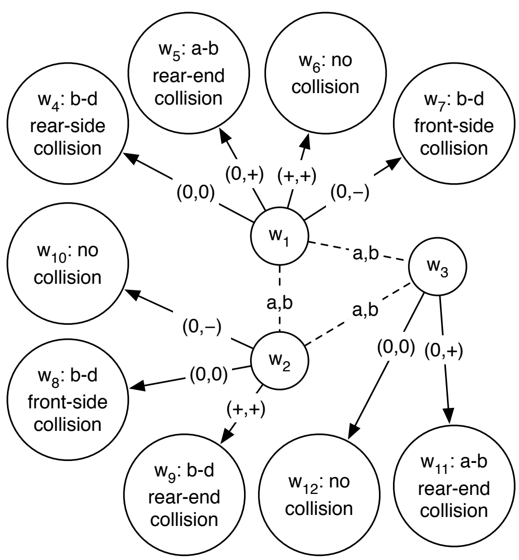

Figure 2: A fragment of an epistemic transition system.

Recall that self-driving car c can observe that car d is not slowing down. Thus, car c can distinguish states w_1 from states w_2 and w_3 of this system. The system might have other states indistinguishable to car c from state w_1. These states, for example, could differ by traffic situations on nearby streets. However, in all these states, coalition $\{a, b\}$ can use strategy $(+,+)$ to avoid a collision. Hence, we say that agent c *knows* how coalition $\{a, b\}$ can avoid a collision. We denote this fact by $w_1 \Vdash \mathsf{H}^{\{a,b\}}_{\{c\}}$("avoid a collision"). In general, we write $w \Vdash \mathsf{H}^D_C\varphi$ if coalition C has distributed knowledge of how coalition D can achieve outcome φ from state w. We call modality H the *second-order know-how modality*. Although in our example coalitions $D = \{a, b\}$ and $C = \{c\}$ are disjoint, we do allow these coalitions to have common elements. Modality H^C_C expresses the existence of a know-how strategy of coalition C known to the coalition itself. Thus, it expresses an existence of a *first-order* know-how strategy of coalition C.

The main goal of this paper is to describe the interplay between the *second-order* know-how modality H and the distributed knowledge modality K. In other words, we axiomatize all properties in the bimodal language that are true in all states of all epistemic transition systems. In addition to the distributed version of S5 axioms for modality K, our logical system contains the Cooperation axiom, introduced by Marc Pauly for strategies in general,

$$\mathsf{H}^{D_1}_{C_1}(\varphi \to \psi) \to (\mathsf{H}^{D_2}_{C_2}\varphi \to \mathsf{H}^{D_1\cup D_2}_{C_1\cup C_2}\psi), \tag{1}$$

where $D_1 \cap D_2 = \varnothing$. Informally, this axiom states that second-order know-how strategies of two disjoint coalitions can be combined to form a single second-order know-how strategy to achieve a common goal. The system also has the Strategic Introspection axiom:

$$\mathsf{H}^D_C\varphi \to \mathsf{K}_C\mathsf{H}^D_C\varphi, \tag{2}$$

which states that if coalition C knows how coalition D can achieve the goal, then coalition C knows that it knows how coalition D can achieve this. In addition, our logical system contains the Empty coalition axiom:

$$\mathsf{K}_\varnothing\varphi \to \mathsf{H}^\varnothing_\varnothing\varphi. \tag{3}$$

This axiom says that if a statement is known to the empty coalition, then the empty coalition has a first-order know-how strategy to achieve it. The axiom is true because the empty coalition can know only statements that are true in each state of the given epistemic transition system. The final and perhaps the most interesting axiom of our logical system is the Knowledge of Unavoidability axiom:

$$\mathsf{K}_A\mathsf{H}^\varnothing_B\varphi \to \mathsf{H}^\varnothing_A\varphi. \tag{4}$$

Formula $\mathsf{H}^\varnothing_B\varphi$ means that coalition B knows that φ will be achieved no matter how agents act. Thus, coalition B knows that φ is unavoidable. The axiom states that if coalition A knows that coalition B knows that φ is unavoidable, then coalition A also knows that φ is unavoidable. To the best of our knowledge, this axiom does not appear in the existing literature. The main technical results of this paper are the soundness and the completeness of the logical system describing the interplay between modalities K and H. The system extends epistemic logic S5 with distributed knowledge by axioms (1), (2), (3), and (4).