Proceedings of IEEE SMC, 2001.

Teams of Agents

GORDON BEAVERS and HENRY HEXMOOR
Computer Science & Computer Engineering Department
University of Arkansas, Fayetteville, AR 72701

Abstract

This paper attempts to focus the discussion about necessary and sufficient conditions for the existence of a "team", the nature of the common mental state required of the agents forming the team, in particular, what an intention is, and various measures of how "team-like" a group of agents is.

1 Introduction

This paper considers fundamental issues about the nature of teams of agents and the mental states of the agents required for the formation of a team. Preceding works in this area have not put forward completely plausible explanations of teams and their formation. The works of Phillip Cohen and Hector Levesque have been very influential and therefore will receive considerable attention in this paper, particularly *On Team Formation* [3]. However, this work is overly restrictive in its requirements for team formation, thereby not counting as teams, groups that should be counted as teams. The notion of *team* developed in this paper will be more inclusive. Michael Wooldridge's recent book *Reasoning About Rational Agents* [15] will also receive attention. Milind Tambe's group has implemented variants of Joint Intentions theories in the "Teamcore" project and are including heterogeneous teams of human and software agents [11]. They consider several of the issues that we will address in this paper such as interagent awareness among team members and joint autonomies.

This paper will address two main questions. First, in attempting to understand what a team is, some criteria need to be established for the existence of a team so that systems of artificial social agents that are capable of teaming might be built that satisfy those criteria. This requires that we have a theory of agents that explains and predicts at least some of the behaviors of agents in social settings. At the very least a "team" is an aggregation of agents with a *common mental state*. This *common mental state* includes having a joint intention, awareness of the team and a cooperative attitude. The second question asks how we are to understand these *common mental states* with respect to artificial agents. Since we are interested in teams of artificial agents and mixed teams of artificial and natural agents, some explication of intentionality and autonomy for artificial agents is required. This paper will advocate that intentions be understood in terms of desires, permissions and abilities.

Addressing the second question seemingly thrusts us directly into the *free will versus determinism* debate, which will not be addressed in this paper. The assumption will be made, therefore, that most humans in most situations have some degree of autonomy and thereby reduce the problem to that of specifying when the internal state of an artificial agent is such that the artificial agent might also be said to have some autonomy. In this paper autonomy will be understood as a combination of ability and permission. In order to avoid the difficult problem of assigning absolute measures to the amount of autonomy possessed by an agent, only relative measures of autonomy will be considered. Autonomy might be measured on a continuum.

The term "team" is used colloquially for a wide variety of aggregations of entities (some of which may be able to have only very low levels of autonomy), from swarm-like groups of ants to groups of fully autonomous individuals sharing a common intention and worldview. Aggregations of agents that may be considered to be teams cover a broad spectrum from the most opportunistic pairs to very formal and large social structures. We follow the common distinction between swarms and teams.

- Swarms are aggregations of entities of very low capacity to form individual intentions. [9]
- Teams are composed of agents that individually have some ability to determine

their own actions in a manner that is not purely reactive.

It is commonly held that swarms of insects do not constitute teams because, although the members of the swarm each have an ability that contributes to the achievement of the swarm's goals through cooperative action, it is questionable that they individually possess the common *mental state* requisite of being team members. Insects most probably do not form intentions, are not aware of being members of a team, and do not consider other team members as beneficiaries of team actions. On the other hand, it is also widely believed that teams of robots are possible with the behavior of the robot being guided by a deterministic program. The behavior of such a robot is thus *innate*. Deterministic systems might be considered to have very little autonomy and thus very little capacity for intentionality. This contrast points out the difficulty of capturing the notions of intention and of autonomy. If fixed responses to given situations are taken to preclude intentionality and autonomy then there will be no teams of robots until the programs determining the behavior of the robots incorporate some method of self-modification, e.g. learning. "Self-determination" does not distinguish between innate and "creative" behavior, but autonomy does.

The notion of autonomy that we are trying to get at here has something to do with complexity, but we will not be able to say exactly what. Ants in a swarm are assumed to be too simple to be able to form intentions although they do individually make decisions that collectively lead to coherent group behavior. Individual ants seem unable to persuade other ants to pursue a goal other than the one adopted. Intentionality implies that agents who are team members have some degree of autonomy. Determining criteria for the existence of a team will require the consideration of concomitant issues such as the importance of a cooperative attitude, abilities, and shared intentions and world view. We also distinguish between a team, which is formed to solve a specific problem, and a coalition, which is a work group formed to maximize the utility of the individual group members (as opposed to the utility of the group itself).

- Members of a coalition share a joint utility and strive for utilitarian stability. [10]
- Members of a team share a joint persistent goal.

Teams are work groups that attempt to achieve a goal even though in doing so they may engage in activities that adversely affect their individual utilities. The team persists so long as the achievement goal persists. Coalitions on the other hand are temporary work groups that pay constant attention to their individual utilities. Stability, viewed as the longevity of the work group, is desirable in coalition formation, however, the sense of stability comes from the expected longevity of positive utilities and there is no notion of explicit persistence as there is when a team is seeking a goal. Other desirable properties of a coalition are efficiency, low cost of forming a coalition, and distributed computational load over the agents joining the coalition. This makes coalitions suitable particularly in economically oriented settings. In comparison, teams seem more suitable when agents must reason about goal achievement and come to an agreement about working together.

Frequently these issues are considered in the framework of a multi-modal temporal logic, e.g., BDI logic. We, however, will take only belief and desire as modal operators and give a definition of intention in terms of autonomy, desire, and choice. We use "state" for what is generally called a "possible world" in Kripke semantics and reserve "possible world" for the branching time structures of temporal logics.

We take an agent's autonomy with respect to an action as a relative measure of its liberty to exercise individual preference with respect to that action. This preference is constrained by social considerations in that it includes consideration of other agents, how the other agents contribute to the agent's sense of freedom to choose and to the agent's performance and how it prefers to work with others. These issues will be combined to yield an assessment of an agent's *permission* to perform an action that will lead to another state. The consequence of an agent's autonomy consideration will determine whether the agent is self-directed, other directed (in the sense of delegating the task to another agent), shares responsibility with other agents (as in teaming), or partially self-directed [4]. Many parameters, both endogenous and exogenous may go into a utilitarian calculation of these preferences [5]. As a more complex preference, an

agent's autonomy determination considers other agents' utilities and social preferences (see [2] for social preferences). Parameters for such preferences in complex agents are hard to enumerate and they change over time with the agent's experience. Only in the most circumscribed situations can we apply normative decision theory to autonomy determination [5]. Although we believe designing parameters is a task that has to be done in each domain, we will show that there are useful domain-independent influences among agent autonomies. Agents may have to revise their autonomy stances in light of the autonomy stances of other agents and might be provoked to assume new autonomy stances.

The personal autonomy (with respect to a given goal) of an individual might vary from complete self-determination (internal) of every aspect of the agent's relationship towards the goal, to complete external determination of those aspects.

In the remainder of this paper we will first develop a formal account of the mental constituents required for teaming. An agent may reflect on these conditions to take steps for bringing about or preserving a team. What these steps might be is beyond the scope of this paper. We will then contrast our approach with that of Cohen and draw some conclusions about teams of agents.

2 Necessary Conditions for the Existence of a Team

In this section we suggest criteria for the existence of a team. In order for a team to exist the agents forming the team must have a shared goal as a part of a shared mental state. We make the simplifying assumption that desires are consistent with beliefs, although this assumption may be relaxed in future work. The terms *belief*, *desire*, and *agent* will be taken as primitive undefined terms. The shared mental state requirement should be understood to include the following four components:

1. The group must have a joint intention, in the sense that there is at least one intention ψ such that each agent of the team intends ψ and believes that the other team members also intend ψ. All intentional agents have autonomy.
2. Each individual team member must believe itself to have some ability that could contribute to the achievement of the shared goal.
3. Each individual team member must be aware of being a team member, and
4. Each individual team member must maintain a cooperative stance toward the members of the team with respect to the shared goal.

Teams are individuated by their joint intention and membership. In order to develop the concept of intention we must first develop the concepts of "ability" and "choice".

2. 1 Ability

Agents must possess an ability (in the sense of a skill), which contributes to the achievement of the team goal in order to be a team member. Unlike humans who may not actually have a capability (in fact) although they believe that they do, we assume that the artificial agents considered here are aware of their abilities. This assumption is made in order to simplify the BDI model. As a result internal and external assessments of the existence of a team should match up. This may pose some problems when mixed teams of artificial and natural agents are considered. [15] gives a definition for first-order ability which we have altered to make it explicit that it is the action α, which might be a complex action composed of many simple actions, which brings about the state ψ. In the case that there are multiple actions which might lead to the same state, an agent may be permitted to perform some of those actions, but not others. It is desirable, therefore, to make explicit the action through which the state can be attained.

Definition 2.1 Can

$$(\mathrm{Can}^0\ i\ \alpha\ \psi) \equiv (\mathrm{Bel}\ i\ ((\mathrm{Agt}\ \alpha\ i) \wedge (\mathrm{Achvs}\ \alpha\ \psi))) \wedge (\mathrm{Agt}\ \alpha\ i) \wedge (\mathrm{Achvs}\ \alpha\ \psi)$$

that is, agent i has a first order ability to achieve the state ψ iff there is an action α that i knows i can perform and i knows that α will achieve ψ. Where (Agt α i) signifies that the agent i carries out the action α, and (Achvs α ψ) signifies that α is the action that leads to the state ψ. We use both *believe* and *know* in this paper without definition and remind the reader that these notion will have to be explicated in a manner consistent with implementation in artificial agents. The above definition leads to a definition for higher order ability, viz.

$$(\mathrm{Can}^k\ i\ \alpha_k\ \psi) \equiv (\mathrm{Can}^{k-1}\ i\ \alpha_{k-1}\ (\mathrm{Can}^0\ i\ \alpha_0\ \psi))$$

for $k > 0$, where α_k may be thought of as a sequence of actions leading to the state ψ.
Generalizing to the multi-agent case yields:

Definition 2.2 Joint-Can

$$(\text{J-Can}^0\ g\ \alpha\ \psi) \equiv (\text{J-Bel}\ g\ (\text{Agts}\ \alpha\ g) \wedge (\text{Achvs}\ \alpha\ \psi)) \wedge (\text{Agts}\ \alpha\ g) \wedge (\text{Achvs}\ \alpha\ \psi)$$

$$(\text{J-Can}^k\ i\ \alpha_k\ \psi) \equiv (\text{J-Can}^{k-1}\ i\ \alpha_{k-1}\ (\text{J-Can}^0\ i\ \alpha_0\ \psi)) \quad \text{for } k > 0$$

In the above formulae α is a rigid designator, that is, α is a specific action that the agent (and each agent in the group) knows will achieve ψ. "Can" stands for the ability of a single agent, whereas "J-Can" stands for the joint ability of a group of agents. See [15] for derivation of "ability simpliciter". Note that the mutual belief required by Wooldridge has been altered to joint belief in the definition of Joint-Can in order to better fit the herding example given below. This also distinguishes our approach from that of Cohen.

Once again, *belief* needs to be explicated with respect to artificial agents but we shall not do so. In J-Can, the agents individually have a belief that they collectively have the ability to achieve the goal and they know of an action that will lead to the achievement of the goal.

In human society we must allow for the existence of teams that are deluded about their abilities, e.g., a team of faith healers with a goal of stamping out cancer in the world is still a team even if the team is completely ineffective. In the context of this definition, however, the agents are taken to capable of an accurate assessment of their abilities. These requirements imply that the agent has the potential to consider the goal and accurately assess the team's possibility of attaining the goal. In order for the agent to have autonomy with respect to the goal, it must furthermore, havepermission to attempt the goal.

The team members believe that they collectively have some ability to accomplish a shared objective, namely, adequate collective ability including know-how. Each individual team member believes that it has some ability that could contribute to the achievement of the shared goal.

2.2 Attitude

In this section we will develop the notions of group autonomy and cooperation. An agent that has the ability to achieve some state ψ may also desire to have this be the case. An agent that has the ability to achieve some state ψ may have social permission to perform an act that will bring the state about. Permissions are derived from social considerations.

Definition 2.4 Autonomy

An agent i is autonomous with respect to a state ψ (through an action α) iff i has the ability to achieve ψ through the action α, but it is not necessary that agent i perform action α, and i is permitted to perform α and also permitted to not perform α, i.e.,

$$(\text{Autonomous}\ i\ \alpha\ \psi) \equiv (\text{Can}\ i\ \alpha\ \psi) \wedge (\text{Can-refrain}\ i\ \alpha) \wedge (\text{Per}\ i\ \alpha) \wedge (\text{Per}\ i\ \neg\ \alpha)$$

where (Can-refrain i α) says that agent i has the ability to refrain from choosing action α. Unlike Can that is rooted at the agent's physical ability, Can-refrain is rooted at the agent's attitude.

Variants of autonomy might include i is permitted to achieve ψ by any means possible or by some means, viz.

$$(\text{AAutonomous}\ i\ \psi) \equiv \forall\ \alpha\ ((\text{Can}\ i\ \alpha\ \psi) \Rightarrow (\text{Autonomous}\ i\ \alpha\ \psi))$$

$$(\text{EAutonomous}\ i\ \psi) \equiv \exists\ \alpha\ ((\text{Can}\ i\ \alpha\ \psi) \wedge (\text{Autonomous}\ i\ \alpha\ \psi))$$

Intuitively speaking AAutonomy might be considered preferable to EAutonomy since the agent's choice of actions is not restricted, but in both cases the agent is considered to have full autonomy. The predicate "Per" stands for the deontic notion of permission for individuals and "G-Per" stands for group permission. See [14] for typical usage of permissions. Note that individuals and groups typically do not have the same permissions and that group permissions cannot be reduced to the permissions of the individuals forming the group.

An agent i may prefer to share its autonomy with a group of agents' g. Agent i would have to believe that, with the other agents of g, it can bring about state ψ, jointly they have liberty (permission) to perform α and that α achieves ψ. This leads to a definition for sharing autonomies between agent members of a group.

Definition 2.5. Group Autonomy

A group of agents g is autonomous with respect to state ψ iff g can jointly achieve ψ through the performance of the action α and has group permission both to do so and to refrain from doing so and it is not necessary that group g perform action α.

$$(\text{J-Autonomous}\ g\ \alpha\ \psi) \equiv$$

$$(\text{J-Can } g\ \alpha\ \psi) \wedge (\text{J-Can } g\ \neg\ \alpha) \wedge (\text{G-Per } g\ \alpha) \wedge (\text{G-Per } i\ \neg\ \alpha)$$

A group's autonomy is relative to a state. We have not elaborated inter-agent autonomies. A reductionist method is being developed elsewhere [5].

Agents with group autonomy could join in group activity. For instance, they could enter a team if they further have a cooperative stance and come to have a joint intention. In addition to shared autonomy, we require agents to have a joint cooperative attitude [12]. We posit that agents must adopt the principle of social rationality in order to be cooperative [7]. Under this principle, agents who are part of a group will prefer actions with the property that the joint benefit to the group is maximized. Social rationality is used to develop a theory of joint responsibility. It is sufficient for social rationality that team members adopt an obligation to cooperate. This implies that team members are jointly responsible for the outcome of this joint activity as well as that each team member considers the other agents as benefactors of shared actions.

Cooperation is the fourth shared mental state required for the existence of a team.

Definition 2.6 Cooperative

An agent is cooperative in a social group g which intends to achieve a state ψ iff the agent believes that it will share in the benefits with the group and it has joint responsibility to achieve that state.

$$(\text{Cooperative } i\ g\ \psi) \equiv (\text{Bel } i\ (\ (\text{Benefits } g \cup \{i\}\ \psi) \wedge (\text{Responsible-for } g \cup \{i\}\ \psi)))$$

"Benefits" is used to denote the group's utilitarian preference for being in a state. "Responsible-for" is used to denote group's willingness to pitch in and carry out any action that they can to bring about the desired state. In general, we consider two varieties of responsibility. The first type is *Responsibility to* which is general, abstract, and typically with respect to an agent's immutable values. This type of responsibility accounts for the agent's high-level qualities that guide its social interaction. We will not consider this further in this paper. *Responsibility for* is specific, concrete, and is typically about an action, a goal, individual autonomy, and commitment. This type of responsibility accounts for inhibiting the agent's low-level actions. It also pertains to normative interactions. We state this more precisely in the following definition.

Definition 2.7 Responsbile-for

A group of agents g is responsible for achieving a state ψ iff any agent in the group who can do an action to bring about ψ will intend to do it.

$$(\text{Responsible-for } g\ \psi) \equiv \forall\, i \in g\ \forall\, \alpha\ ((\text{Achvs } \alpha\ \psi) \wedge (\text{Can } i\ \alpha\ \psi)) \Rightarrow (\text{Int } i\ \alpha\ \psi)$$

Definition 2.8 Joint Cooperation

A group g of agents is jointly cooperative with respect to state ψ iff each agent of the group g is individually cooperative toward that state

$$(\text{J-Cooperative } g\ \psi) \equiv \forall\, i \in g\ (\text{Cooperative } i\ g\ \psi)$$

2.3 Intentions

We assume that when an agent adopts an intention, the agent does so only after consideration of its desires and its autonomy with respect to those desires. Desires are possible worlds which we assume the agent is able to partially order according to desirability (utility). We defer the discussion of the conditions that need to hold in order for this ordering to be ascertained in a reasonable amount of time. The ordering may have to be approximate in order to avoid excessive computational complexity. Using this ordering the agent applies a choice function to select that action/state combination with the highest utility. The result is that an action is chosen by the agent to cause a transition to the desired state. We use the notation (Choice i α) to indicate that the application of the choice function by the agent i yields the action α. The state at which the choice function is implemented will affect the result, but the notion hides this component. Note that it is immediate from the definition of *Individual Intention* that intentionality implies autonomy, which is in accordance with common sense.

Definition 2.9 Individual Intention

$$(\text{Int } i\ \alpha\ \psi) = (\text{Autonomous } i\ \alpha\ \psi) \wedge (\text{Choice } i\ \alpha)$$

2.4 Joint Intentions

The agents forming the team must individually intend that the team achieve the goal. The above definition interprets the notion of intention for artificial agents, but we leave vague what it means for a natural agent to intend an action, and reiterate that there are complications introduced by allowing agents of disparate *mental capacities* to team to-

gether. In particular, what does it mean for a human and a robot to intend the same action?

Teaming requires coordinated action of the team members and this coordination in turn requires that the individual agents be able to determine when to start or stop an activity, when to switch activities, and when to adjust their contribution. A personal computer coordinates its activity with the activity of its user, e.g., echoing keystrokes or implementing a screen saver but does not constitute a team together with its user because the computer lacks the ability to form intentions. The computer's responses are hard wired. A potential team member thus must have the ability to accept or reject a goal, that is, form intentions. Part of this ability is the autonomy to make this decision, that is, without the autonomy to adopt an individual intention an agent cannot become a team member.

Commitment to the intention (goal) is part of being a team member and this commitment will demonstrate itself in the persistence of the intention. Cohen and Levesque ([3]) have investigated this issue. Joint intention will be defined in the pattern of individual intention.

Definition 2.10 Joint Intention:

$$(\text{J-Int } g\ \alpha\ \psi) \equiv (\text{J-Autonomous } g\ \alpha\ \psi) \wedge (\text{Choice } g\ \alpha)$$

2.5 Awarenesss

Our notion of awareness captures an agent's cognizance of its team-mates' actions. This cognizance may rely either on the agent's plan recognition abilities, communicated messages, or direct observations. In this paper we will not expand on varieties of means for becoming aware nor do we give an account of an ongoing awareness maintenance (or active monitoring) as in [8]. Instead we will capture this awareness in an agent's belief about actions of another agent.

A basic requirement for the existence of a team is coordinated action, and it is impossible for an agent to coordinate when she does not perceive what she is to coordinate with. Each individual team member must be aware of being a team member, i.e., some ability to infer that other agents are on the team. The agents forming the team must individually be aware of the existence of the team and intend to coordinate their actions with other perceived team members. Since teams are individuated by intention and membership, the agent must know what the joint intention is and that it is part of the team with that intention.

Definition 2.11 Belief-Aware

An agent is belief-aware of a social group g that intends to achieve a state ψ iff the agent believes that together with the group it has joint ability toward the desired state and that it has joint intention to achieve that state

$$(\text{B-Aware } i\ g\ \psi) \equiv (\text{Bel } i\ ((\text{J-Can } g \cup \{i\}\ \psi) \wedge (\text{J-Int } g \cup \{i\} \psi)))$$

The members of the team have to be jointly aware of the other team members' intentions to contribute.

Definition 2.12 Joint Belief-Aware

A group g of agents is jointly belief-aware with respect to state ψ iff each agent of the group g is individually belief-aware of that state

$$(\text{J-B-Aware } g\ \psi) \equiv \forall i \in g\ (\text{B-Aware } i\ g\ \psi)$$

2.6 Putting it altogether

We are now ready to compose the requirements we have outlined into one definition that captures the notion of a team.

Definition 2.13 Team:

A group of agents g is a team iff they have joint intention, group autonomy, joint belief-awareness, and Joint cooperative attitude.

$$(\text{Team } g\ \psi) \equiv (\text{J-Intention } g\ \psi) \wedge (\text{G-Autonomy } g\ \psi) \wedge (\text{J-B-Aware } g\ \psi) \wedge (\text{J-Cooperative } g\ \psi)$$

3 Discussion of Related work

"On Team Formation" by Cohen, Levesque and Smith [3] has been an influential paper. Comparing our views to theirs will help clarify our position with respect to team formation. Cohen et al claim that speech acts are required for team formation. The primary focus of "On Team Formation" is Cohen's attempt to link joint action theory with speech act theory although *speech acts* are interpreted rather broadly, e.g. "significant look". Although speech acts are in fact commonly used in forming and disbanding teams of humans, there are examples of teaming in which there is no identifiable speech act. For example Jill comes across Jack attempting to herd a flock of sheep, which is too

large for him to control. Jill spontaneously and without communication of any kind begins to cover one flank and Jack, taking notice of Jill's efforts, concentrates his attention on the other flank. This is an example of coordinated agents with a shared goal and, in a loose sense, shared mental state, that is, a team. The pair satisfies the four requirements given in the previous sections for the existence of a team, "One might say that they are working together as a single unit. They are not a coalition since they have an explicit goal in mind and agreed implicitly. This implicit agreement is achieved on the basis of situational norms.

Notice that the example would be essentially the same if one of the pair was human and the other canine, or if both were canine, or if both were robots, or any other combination provided that the four criteria can be attributed to the agents. In the following we shall refer to this as the "herding example". This example has guided our development of the criteria for the existence of a team in the preceding sections. This example calls into question Cohen's claim that there is a "tight relationship" between speech acts and the formation of teams. The linkage between speech acts and joint action theory provided by "On Team Formation" should be viewed as illustrative of a common procedure for the formation of *formal* teams among polite humans following social conventions, but by no means, exhaustive of the possibilities. We use *formal* to mean some contract like process has been followed, e.g. a request followed by an acceptance as specified by Cohen.

Mere coordination of activity is insufficient to constitute the existence of a team. Drivers on a public road coordinate their actions for their mutual benefit and to facilitate the achievement of their individual goals. They share the goal of avoiding physical contact with one another. Cohen claims that "(t)he key property distinguishing joint or collaborative action from mere coordinated action is the joint mental state of the participants." They share the mental state of being attentive to their driving and jointly having the intention to avoid contact with one another, but this is not sufficient to make them a team. The crucial distinction is between having a "joint intention" (a token) and jointly having an intention (a type). The shared goal and shared mental state of these drivers is insufficient to make them a team since there is no joint intention. The problem then is to explicate *shared goal* and *shared mental state* in such a way that these drivers are not a team but two robots herding together are a team. We, and Cohen, claim that the shared goal of the drivers to avoid contact is too nebulous to bind the drivers together in a team, but the question then becomes "how precise must the goal be in order to contribute to the existence of a team?" "On Team Formation" does not address this question.

Cohen claims, "[t]he best way to explore what this mental state must be is to imagine a joint action going astray. Our favorite example is driving in a convoy, versus ordinary traffic. If one driver observes another leaving the road, there is no requirement that the observer take any specific action (except perhaps, avoidance). However, if the drivers are in convoy, and one driver is observed pulling off the road, the other drivers will do likewise. ([3], page 1)"

Cohen asserts that the contribution of [3] is its having provided an analysis of the communicative actions that form and disband teams. The implication is that communicative actions are required for forming and disbanding teams, which the herding example above contests. We agree with Cohen, however, that a *joint intention*, i.e., "a joint commitment to perform a collective action while in a certain shared mental state" is a fundamental requirement for the existence of a team.

Cohen claims that because the agents forming a team may hold different beliefs, communication becomes necessary and indeed that "special communicative demands are placed on agents involved in joint activities ..." ([3] page 3). The implication is that agents holding different beliefs relative to the task at hand cannot be a team; the team members must have *mutual* beliefs and common goals in order to constitute a team, and the only causal mechanism to accomplish this isomorphism of mental states is communication. Consideration of the shepherd and her dog calls this claim into question. Although they form an effective herding team and some communication between them is typical but not necessary, this communication does not result in the isomorphism of higher order beliefs that Cohen seems to think necessary for the existence of a team, and furthermore, the dog most probably is not capable of forming the higher order beliefs required for mutuality.

Cohen claims that "(t)he essence of team behavior is cooperation." ([3] page 4) and contends that team behavior requires joint commitment to each other's success and the success of the overall project. These assumptions seem unassailable given a broad enough interpretation of *joint commitment*.

Key to the understanding of shared mental state is the understanding of *joint intention*. "On Team Formation" provides the following as a first attempt to define *joint intention*: "x and y jointly intend to do some action iff it is mutually known between x and y that they each intend that the collective action occur, and also that they each intend to do their share as long as the other does likewise, and this mutual knowledge persists until it is mutually known that the activity is over (i.e., is successful, unachievable, irrelevant). ([3] page 4, modified page 12)"

Cohen recognizes that this definition is too strong "because it stipulates at the outset that the agents must mutually believe that they will each have their respective intentions until it is mutually known that they do not." ([3], page 5) Cohen later modifies the definition to rectify this shortcoming, but the key element, *mutual knowledge* remains in the revised definition. The motivation for Cohen's concern is "doubt-induced unraveling of the team effort," that is, what happens to the team when one or more members come to believe that the goal is unachievable.

One test of the adequacy of Cohen's definition of joint intention will be whether it distinguishes between the mental states of drivers and herders. Suppose driver x and driver y each intends to contribute to the smooth flow of traffic and they mutually know that the other also has this mutual intention, and further that they mutually intend to continue to do their share. Knowledge is taken to be mutual true belief. These drivers have a common goal and a joint intention under the above definition. The joint intention is a shared mental state so they are thus a team by the standards provided by Cohen, contrary to the widely held belief that they are not a team. Thus the suggested definition of joint intention is too broad unless the mutuality of the knowledge is interpreted in some narrow way, e.g. as requiring communication, but then it would be too strong as it would then exclude the herding example of teamwork.

We doubt that mutual knowledge of intentions is the central concept underlying joint intention, rather we contend that the essence of a team is the cognizance that individual members of the team take of one another's actions and of the overall situation that the team finds itself in, and furthermore the intention of each of the team members to cooperate, that is, to respond to the changing situation and to one another's action in a manner appropriate for the attainment of the joint intention. Under this interpretation teams can be extremely volatile, forming and disbanding with little or no effort. Consider the situation where one driver spontaneously teams with another just long enough to clear a traffic jam, then, their goal achieved, the team behavior is abandoned. The team has been formed and disbanded without any spoken communication.

4 Conclusion

We have developed a formal account of a team that relies on some joint understanding of a set of mental notions: intentions, autonomy, awareness, and benefit. Our account remedies shortcomings with the popular joint intentions theory of [3]. Particularly, nontrivial abilities of team members, interagent awareness, and benefit were introduced as additional criteria for a team. In our formulation, we have reduced intention to desire and belief and also uncovered the relationship between mentalistic notions of BDI and decision-theoretic accounts for teaming, which we plan to explore further in future work.

Acknowledgements

We thank Chris Hill and the anonymous reviewers for their helpful comments. This work is supported by AFOSR grant F49620-00-1-0302.

References

[1] S. Brainov and H. Hexmoor, 2001. "Quantifying Relative Autonomy" in **Multiagent Interaction, In IJCAI-01 Workshop, Autonomy, Delegation, and Control.**

[2] S. Brainov, T. Sandholm, 1999. "Power, Dependence and Stability in Multiagent" in **Plans. AAAI/IAAI 1999**: 11-16.

[3] P. Cohen, H. Levesque, I. Smith, 1997. "On Team Formation" in J. Hintika and R. Tuomela, **Contemporary Action Theory**, Synthese.
[4] H. Hexmoor, 2000. "A Cognitive Model of Situated Autonomy" in Proceedings of **PRICAI-2000** Workshop on Teams with Adjustable Autonomy, Australia.
[5] H. Hexmoor, 2000. "Case Studies of Autonomy" in proceedings of **FLAIRS 2000**, J. Etherege and B. Manaris, eds. pp. 246- 249, Orlando, FL.
[6] H. Hexmoor, 2001. "Stages of Autonomy Determination" in preparation.
[7] N. R. Jennings and J. R. Campos. "Towards a Social Level Characterisation of Socially Responsible Agents" in **IEE Proceedings on Software Engineering**, 144 (1), 1997, 11-25.
[8] G. Kaminka, D. Pyandeth, M. Tambe. "Monitoring Deployed Agent Teams", International Conference on Autonomous Agents (**Agents '01**).
[9] J. Kennedy and R. Eberhart, **Swarm Intelligence**, Morgan Kaufmann, 2001.
[10] O. Shehory, K. Sycara, and S. Jha. "Multi-agent coordination through coalition formation" in M. P. Singh, A. S. Rao, and M. J. Wooldridge, eds.**, Proceedings of ATAL'97**. Springer Verlag, Heidelberg, Germany, 1998.
[11] M. Tambe, D. Pynadath, C. Chauvat, A. Das, and G. Kaminka, 2000. "Adaptive agent architectures for heterogeneous team members" in **Proceedings of the International Conference on Multi-agent Systems (ICMAS 2000**), Boston, MA.
[12] R. Tuomela, 2000**. Cooperation: A Philosophical Study, Philosophical Studies Series,** Kluwer Academic Publishers.
[13] E.M. Verharen and F. Dignum. "Cooperative information agents and communication" in M. Klusch, ed., **Proc. of the 1st Int.l. WS on CIAs**, Berlin, 1997. Springer-Verlag.
[14] R. J. Wieringa and J.-J. Ch. Meyer, 1993. "Applications of Deontic Logic in Computer Science: A concise view", in **Deontic Logic in Computer Science**, 17 – 40, John Wiley & Sons, Chichester, England, (1993).
[15] M. Wooldridge, 2000. **Reasoning about Rational Agents**, The MIT Press.