Forming Shared Mental Models

Katia P. Sycara
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
katia@cs.cmu.edu

C. Michael Lewis
Department of Information Science
University of Pittsburgh
Pittsburgh, PA. 15260
ml@pigeon.lis.pitt.edu

Abstract

As problems increase in complexity it becomes impossible for any one person to know all the things necessary to make good decisions. A group of specialists, even if they possess the requisite knowledge, remain a collection of individuals until their expertise can be jointly brought to bear. The problem of fusing expertise where individuals have very detailed knowledge in their own areas and much weaker understanding of others is that no one knows what anyone else needs to know. This impasse cannot be broken until shared mental models are developed to provide the common perception needed to focus the activity of the group. This paper presents characteristics of shared mental models and a model of the effect, nature, and process of the formation of shared mental models in cooperative problem solving by a team of specialists.[1]

Introduction

Decision making in a variety of complex tasks requires the cooperation of many decision makers. Very often, and when the task is especially complex, each decision maker is also a specialist in a particular area. At the same time decision making in a team of specialists is frought with difficulties. A major difficulty is that different specialists lack (1) a shared language for communication and (2) shared perceptions of the task. Evidence supporting this assumption comes from a variety of sources. Case studies of decision making in organizations (e.g., Bond's [Bond 89] Lockheed study of aircraft design) have found that specialists do not understand the details of each other's models and language, but through cooperation and interaction are somehow able to produce designs of very complex artifacts, such as aircraft. Research in group decision making has found that the variability of approaches to decision making across groups appears to be even greater than across individuals [DeSanctis 87]. The only consistently reported difference between successful and unsuccessful problem-solving groups has been that successful groups devote adequate time to *problem formulation* and *planning of meeting strategy*, whereas unsuccessful groups immediately begin to search for alternative solutions [Hirokawa 83, Hackman.Kaplan 74]. In other words, successful groups spend adequate time to build shared vocabulary and mental models of the nature of the decision problem, strategies, significance of information, and participants' roles. Shared mental models have also been hypothesized by [Athans 82] and [Fischhoff 86]. The importance of the formation of shared mental models has also been experimentally observed in time critical high-stress decisions such as the air crew emergencies studied by [Orasanu 90].

Although the evidence exists that the formation of shared mental models is a major factor in successful group decision making by humans, little has been done to characterize what "shared mental models" may be, the processes by which they are formed, or their role in group decision making.

Similarly, Distributed AI research sheds little light on this problem. Most work in DAI and team decision making has focused on homogeneous agents [Durfee 87, Cammarata 83, Rosenschein 85, Sullivan 89]. In such situations, the agents already share a common perception of the task and based on this understanding they have to coordinate their *actions*. If they communicate at all, communication of high level agent plans [Durfee 87], based on which agents form expectations of other agents' behavior, ensure coordination. In the team of specialists situation, the agents are by definition heterogeneous, they do not share a common perception of the problem, or each others' expertise. Yet they have to make highly interdependent decisions with uncertain outcomes. A prerequisite to coordinating their actions is to coordinate their *thoughts* through the formation of shared mental models.

In particular, we present the following characteristics of shared models:

- Shared models act as a basis for inferring opportunities for cooperation

[1] This research has been supported by DARPA and AFOSR under contract number F49620-90-C-0003.

- Shared models form the basis for identifying another's needs resulting in helpful intervention
- Shared models act as a basis for inferring another agent's capabilities (resulting in the agent asking for help) and limitations and taking them into consideration in communicating with other agents
- Shared models serve as a basis for knowing *what* to communicate (i.e. what will be understood)
- Shared models act as basis for conflict resolution
- Shared models act as basis for optimal solution integration

The formation of shared mental models is an incremental process consisting primarily of information communication and successive updating and refinement of a sketchy understanding of the ramifications of the problem that the team is solving. In this paper, we present a model of agent models that are active during group decision making and their interactions to forming shared mental models in problem solving by a team of specialists. The domain of our study is concurrent engineering design [Sycara 89a]. The model is currently being implemented in CADET, a design problem solving system [Sycara 90].

Cooperative Problem Solving by a Team of Specialists

The present model of decision making by a team of specialists extends earlier work [Sycara 89b, Sycara 90] characterizing group decision making as a negotiation process. A team of specialists is considered to be a group of individuals with common goals, each with highly specialized knowledge in a particular area but with less precise knowledge of other areas. A normative group decision, therefore, is one which integrates the relevant portions of this specialized knowledge within a common model. The key to this definition lies in the determination of relevance. We presume that in a group decision process of this sort, it is neither feasible nor desirable to form a common model incorporating all of the group's expertise. This would require making M.D.'s of physicists and engineers of accountants. Instead, a normative group decision is one generated by a negotiation process which efficiently elicits some small subset of the group's expertise which determines a decision as good or better than other possible negotiations of comparable length.

Group decision making by cooperating specialists can be viewed as a multi-agent planning task where each agent has (a) incomplete knowledge of the environment, (b) limited knowledge of the constraints and intentions of other agents, and (c) limited number and amount of resources that are required to produce a system solution. When specialists cooperate, they bring together multiple viewpoints and diverse knowledge on a single problem. Bringing together diverse knowledge is a source of robustness and balance. The team can solve problems that are beyond the scope of any of the individual members. Furthermore, the solutions are generated from a rich and varied body of knowledge which provides potentially, a bigger set of good solutions to choose from and the potential for creativity. On the other hand, there are difficulties with resolving the conflicts that arise when trying to merge multiple goals, priorities, and evaluation criteria that are the results of individual expertise. Thus, each agent has a limited and egocentric view of the problem. This may result in misunderstandings, goal conflicts and solution suboptimalities. Because of the lack of appropriate expertise in all areas needed to solve the problem, and because of the presence of conflicting constraints, goals and possibly evaluation criteria, it is impossible for each agent/expert/specialist[2] to reach an optimal solution using only local information. Typically the decisions of one agent impact the decisions of another and vice versa. Thus, we cannot simply model each agent. Rather we must augment an agent's problem solving process by including yet another model, the "shared model" that captures his interactions and decision-coordination with the other agents.

Decision making by a team of specialists has the following characteristics:

- The global system goal is to produce a decision that is synthesized from contributions of different expertise, concerns and constraints.
- During the process, conflicts in the form of constraint and goal violations could arise.
- Disparate evaluations of a situation could surface as a result of different criteria used to evaluate the situation from different perspectives. Typically, these criteria cannot be simultaneously and optimally satisfied.
- Another kind of conflict is over approaches used to achieve a goal.
- The system goal is achieved by making the best tradeoffs on conflicting goals and constraints.

As a result of these characteristics, there must be a process that helps incrementally incorporate in the final decision the various types of available expertise, and the best tradeoffs that could be made in a given situation.

The dilemma confronting a team of experts is that the subset of individually held expertise that is relevant to solving the problem at hand cannot be known in advance. Each expert has detailed knowledge of some aspect of the task and beliefs, based in part on inexpert understanding of other aspects, about how this expertise may be related to the groups goals. Because no member knows precisely

[2]We use the words "agent", "expert" and "specialist" interchangeably for the purposes of this paper.

what any other knows, none can determine the relevance of their own expertise, if that relevance depends on things known only by others. The situation is similar to that of a group attempting to assemble a complex puzzle. Each member has some pieces mixed in with pieces from a great many other puzzles. Players can only view their own pieces and there are costs associated with making pieces public by placing them on the table. If players hold pieces without prominent identifiers such as matching background color, the relevance of these pieces can only be established through an iterative process of advancing and retracting pieces until a context is developed allowing the players to "see" which ones fit.

In the design domain for example, a tricycle designer knows (from his expert knowledge) that using cotter pins and caps to hold the rear wheels and pedals of a tricycle together results in much higher performance (leading to higher quality) than using press-on caps. Because of his ignorance of costs, and his approximate knowledge that quality is linear with costs, the designer assigns a higher cost to a design that uses cotter-pins than to one that uses press-ons, thus making him indifferent to the choice. Similarly, the manufacturing engineer knows (from his expert knowledge) that drilling round stock to make the hole for the cotter pins is a more expensive process than fitting the press-on caps, so (having the inaccurate general knowledge that quality linear with cost), he is also indifferent to the choice. The union of their expertise, however, may strongly favor one or another of these alternatives. For instance, if the designer rated the cotter pin design as four times more desirable than the press on caps and the manufacturing engineer rated the cost at less than double, the cotter pin design is favored. This representation of the compartmentalization of expertise captures the features we believe are essential to describing the pooling of expertise. Each expert has specialized knowledge about his part of the problem but misconceptions about other agents' parts. As a consequence, no expert knows which parts of his expertise are really needed by other members of the group.

We claim that expert team decision making solves this problem through iteratively updating these naive models via the communication of the group members' expertise resulting in the development of "shared models".

Models of Agents

The model we are proposing operationalizes private and shared knowledge in the form of expert domain knowledge, utility functions and constraints and the formation of shared models as a negotiation [Sycara.aor, Sycara.eurjor90] modifying these structures. The model is quite general and independent of decision protocols, describing with equal ease consensus, voting, or advisory regimes. The model provides normative references for group decisions by assessing product and process simultaneously. A group decision can be judged either with respect to the publically available model at the time of decision of the distribution of normative decisions for possible models resulting from negotiations of equivalent or shorter length. The assumption that group decision making is a satisficing process trading off time and effort for decision "quality" is a central tenet of this approach.

In complex domains, such as design, attributes of concern (e.g., design attributes such as material out of which an artifact is made) interact in complex non-linear ways with the objective to be maximized, profits. This interaction happens through mediating variables, such as functionality, durability, aesthetic attractiveness etc. The question that arises is, what is suitable information aggregation for inclusion in the "shared models"? How does a "shared model" interact with individual specialized knowledge? What is a suitable representation of individual expertise and "shared models"? What are the variables used for communication among group members?

Our hypothesis is that each agent:

- Has a model of his individual unique expertise, called the *"expert model"*, characterized by detailed knowledge about some particular aspect of the task.
- Has a naive understanding of aspects of the problem outside of his area of expertise, called the *"naive model"*. The naive model characterizes weak commonly held beliefs such as, "the more expensive a material is, the more durable it will be."
- Develops through interaction with other specialists a more comprehensive model of the problem at hand, called the *"shared model"*, which incorporates elements of others' expertise. The "shared model" between two agents also defines the common vocabulary that the two agents can use to communicate in an intelligible way.

In general, "shared models" might not be the same across all agents. For the group to make a common decision, the potentially different"shared models" that have been formed through interactions of subsets of agents must converge to a common view of the global problem which captures the important variables and decisions that must be coordinated. The parts of different agents' "shared models" that are common across the group comprise the *"common (group) model"*.

During decision making, agent models appropriate for the task are composed by substituting the corresponding parts of an expert model for the naive model. The utility an agent associates with a particular alternative, therefore, reflects both components. As group problem solving progresses, agent models are modified to incorporate expertise imparted by other agents. These modifications lead to the incremental building of shared models among sets of agents.

Model Characteristics

Models of agents are hierarchically organized in the form of increasingly refined domain concepts, constraints and utility functions. The complex representation at the bottom of this hierarchy corresponds to the full expertise of the agent which may or may not be needed by the group for a particular decision. At the top of the hierarchy there is sketchy knowledge, the naive models shared by the members of the group. The naive models are required to be accurate in aggregating across situations but may be imprecise in capturing distinctions among situations made by the expert model. In other words, using only his specialized knowledge and the naive model, each specialist would be led to erroneous decisions. Naive model imprecisions may arise through naive approximation of the values, relations and constraints among the attributes of the problem variables.

The mental models we are investigating have a variety of characteristics. At the beginning of problem solving, an agent's mental model consists of his expertise and a naive understanding of other specialties which is refined in the course of the consultation developing into a (set of) shared models. The naive model contains aggregate variables that constitute the vocabulary through which enables the different specialists to begin the problem solving interaction.

An agent's expertise consists of facts, constraints, their relations and utilities that lie in his specialty. For example, a designer's expertise contains the constraints among design attributes such as artifact components, component materials, types of allowable structural connections among components, and artifact dimensions. Utilities are associated with particular values of these artifact-descriptive variables that operate in the design space. Utilities are in the range of [0, 1]. For example, the utility of making a tricycle out of light steel may be .8, making it out of heavy steel .6, making it out of wood 0 and so on.

The naive model consists of a set of relations among aggregate variables. In design, these relations relate the quantities profits, design quality, cost, price and saleability. These relations are at an aggregate level and are therefore only approximately accurate. For example, the naive model may say that production quality is proportional to production costs, unit sales inversely proportional to price, and willingness to pay (saleability) (for the decreasing number who could) proportional to quality. As a result, the naive model predicts that saleability will be independent of design decisions balancing cost and quality. (The group's profit margin can be maximized by choosing the design which minimizes the cost while maximizing the quality). In other words the naive model is indifferent among design options. This evaluation made by the naive model is not accurate since it does not take into consideration precise relations between variables of the problem. The naive model constitutes a common and unbiased basis which is transformed and refined through the group interaction.

The naive model and an agent's expertise are connected through a set of relations that map variables in the *expert space* to the variables in the *naive space*. The mappings contain both accurate and approximate knowledge. The accurate knowledge is the set of relationships that map variables in the expert space to *mediating* aggregate variables that are within the area of agent expertise. For example, the designer's expert space variables, such as the artifact-specific attributes, their constraints, utilities and interrelationships are mapped into the inexpert space that contains the variables profits, cost, price and quality through mediating variables, such as durability, structural soundness etc. For the designer, these intermediate variables map to design *quality* through a set of relationships. Design quality is an aggregate variable in the domain of expertise of the designer that is also present in the naive model. The set of intermediate relationships that map the variables within an expert's specialty to the naive model is available only to the particular expert and is not used for intra-group communication. On the other hand, relations between aggregate variables, within and outside the domain of expertise of a specialist are incompletely known to him. For example the relation between quality (an aggregate variable within the domain of a designer's expertise) and cost is only approximately known to the designer, since cost is not one of the aggregate variables in his domain of expertise. This interaction of accurate and inaccurate knowledge may lead the expert to incorrect inferences. Expert team decision making solves this problem through iteratively updating the naive models via the communication of the group members' expertise resulting in the development of a shared model.

An Example

Consider the decision situation for a team of specialists in a manufacturing enterprise tasked with concurrently engineering the design of a tricycle. The team objective is to optimize tricycle design. A design will be optimal if it maximizes profits, under certain assumptions relating design attributes to cost, price, and ease of selling the tricycle.

The design attributes have associated utilities that express the desirability of particular attribute values with respect to different evaluation criteria. For example, in terms of tricycle durability, high grade plastic has utility .8, whereas low grade plastic .4. In terms of cost, high grade plastic has utility .6 whereas low grade .8 (high grade plastic is more expensive than low grade). Each team member has expertise in terms of optimizing only part of the design (e.g., cost of material, ease of manufacturing, reliability). Because of the inherent tradeoffs of attribute values with respect to different criteria not all parts of the design can be simultaneously optimized. Furthermore, the overall evaluation function that incorporates cost tradeoffs is not precisely known. In a group decision making situation, each expert knows the relation among attributes in his domain of expertise and the payoff associated with each attribute value alternative, but he is unaware of (a) the relations among attributes in someone else's domain of expertise, or (b) how the attributes in his domain of expertise impact the overall cost function. For example, the designer knows that in terms of strength of the tricycle frame, braced and welded frame has the highest payoff, bolted the next highest, and integral the lowest. In terms of weight, high grade plastic has the highest payoff, with heavy steel tubing the lowest.

The group's goal is to maximize profit and may be expressed as:

profit = unit_sales x (unit_price - unit_cost)

A negotiation over this decision might proceed as:

Manufacturing Agent: Do we want to use press-on caps or cotters for the wheels and pedals The drill press operations will add another 3% to manufacturing costs. (This is an informational exchange. The manufacturing agent is indifferent to this decision because it does not affect the group's goal as predicted by his agent model.)

Design Agent: In that case, I think we should use cotters. The press-on caps are likely to start falling off after 6 months-1 year while the cotter pins will hold the wheels on indefinitely. (The reliability of the artifact which supports the fulfillement of the public attribute of quality is increased by 500% at a manufacturing cost of only 3%, therefore by the naive portions of the designer's model relating quality to profits through unit sales this alternative should more than double profits furthering the group's goal.)

Sales Agent: I don't think we should do it. A buyer can't see that there is a cotter under the cap and therefore it has no effect on perceived quality. (Perceived quality is the sale's experts private version of the public attribute quality and is the attribute in the group's expert model which relates quality to unit sales. It therefore takes precedence.)

All three agents now agree because with this modification, their shared model predicts that the increased production cost associated with cotters will be detrimental to the goal of increasing profits because there is no offsetting influence of this form of quality on unit sales.

Concluding Remarks

The model proposed in this paper is intended as a research tool for the study of group decision making by people and machines. The model provides the first normative treatment of group decisions which incorporates the very features motivating this form of decision making. The importance of shared mental models has been recognized in the cognitive decision making literature [Athans 82, Fischhoff 86] but no investigation as to the process of forming shared mental models has been proposed to date. We have proposed the existence of a variety of mental models (the expert model, the naive model, shared models and a potentially common model) during group decision making and presented their interactions during the problem solving process. In addition, we have presented a concrete approach to describing those models and the process by which they are formed. The shared mental models discussed in this paper define the nature and minimal level of aggregation of the information necessary to be common knowledge to agents so that suitably coordinated decisions can be made. It is our hope that this model can provide a tool for the study of bias in group decision making which may prove as fruitful as normative models have been in the study of individual decisions.

References

[Athans 82] Athans, M.
The Expert Team of Experts Approach to Command-and-Control Organizations.
IEEE Control Systems Magazine September:30-38, 1982.

[Bond 89] Bond, A., and Ricci, R.
Cooperation in Aircraft Design.
In *Proceedings of the MIT-JSME Workshop on Cooperative Product Development*. Cambridge, Mass., 1989.

[Cammarata 83] Cammarata, S. McArthur, D. and Steeb, R.
Strategies of cooperation in distributed problem solving.
In *Proceedings of IJCAI-83*, pages 767-770. IJCAI, Karlsruhe, W. Germany, 1983.

[DeSanctis 87] DeSanctis, G., and Gallupe R. B. .
A Foundation for the Study of Group Decision Support Systems.
Management Science 33:589-609, 1987.

[Durfee 87] Durfee, E.H.
A Unified Approach to Dynamic Coordination: Plannign Actions and Interactions in a Distributed Problem Solving Network.
PhD thesis, COINS, University of Massachusetts, 1987.

[Fischhoff 86] Fischhoff, B.
Decision making in complex systems.
In Hollnagel (editor), *Intelligent Decision Support in Process Environments.* Springer Verlag, New York, N.Y., 1986.

[Hackman.Kaplan 74] Hackman, J.R. and Kaplan, R.E.
Interventions into Group Process: An Approach to Improving the Effectiveness of Groups.
Decision Science 5:459-480, 1974.

[Hirokawa 83] Hirokawa, R.Y., and Pace, R.A.
A DEscriptive Investigation of the possible Communication-based Reasons for Effective and Ineffective Group Decision Making.
Communication Monographs 50:363-379, 1983.

[Orasanu 90] Orasanu, J.
Shared Mental Models and Crew Performance.
In *Proceedings of the 34 Annual Meeting of the Human Factors Society.* Orlando, Fla., 1990.

[Rosenschein 85] Rosenschein, J.S.
Rational Interaction: Cooperation Among Intelligent Agents.
PhD thesis, Stanford University, 1985.

[Sullivan 89] Sullivan, D., and Kleinman, D.L.
A Normative Descriptive Approach to Team Distributed Scheduling.
In *Proceedings of the 1989 IEEE International Conference on Systems, Man and Cybernetics.* IEEE, Cambridge, Mass., 1989.

[Sycara 89a] Sycara, K. and Navinchandra D.
A Process Model of Experience-Based Design.
In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society.* Ann Arbor, MI., 1989.

[Sycara 89b] Sycara, K.
Argumentation: Planning Other Agents' Plans.
In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89).* Detroit, Mich, 1989.

[Sycara 90] Sycara, K.
Viewing Design as a Cooperative Task.
In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society.* Boston, Mass., 1990.