



Faculty Publications

1993-10-01

Coordinated intelligent control via epistemic utility theory

Wynn C. Stirling
wynn_stirling@byu.edu

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Electrical and Computer Engineering Commons](#)

Original Publication Citation

Stirling, W. C. "Coordinated Intelligent Control Via Epistemic Utility Theory." *Control Systems Magazine*, IEEE 13.5 (1993): 21-9

BYU ScholarsArchive Citation

Stirling, Wynn C., "Coordinated intelligent control via epistemic utility theory" (1993). *Faculty Publications*. 705.

<https://scholarsarchive.byu.edu/facpub/705>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Coordinated Intelligent Control Via Epistemic Utility Theory

Wynn C. Stirling

Epistemic utility theory is used to model autonomous intelligent agents. A *epistemic system* provides a mechanism for agents to characterize their knowledge corpora, options, goals, and beliefs. Levi's rule of epistemic utility provides a principle of action for decision making by comparing the informational value of rejection with the belief of correctness. Decisions are made locally and reactively, rather than globally. Coordination is implemented between agents by sharing and learning the epistemic systems of other agents. The resulting coordination model is nonhierarchical, heterogeneous, and does not require explicit communication between agents.

Multiple Agent Coordination

Intelligent control requires agents to act autonomously, and to do so in a manner that permits each member of a multiple agent group to coordinate its behavior with other agents in its environment. Multiple agent coordination has many facets, ranging from idealized situations where instantaneous communication between agents is possible and all agents are committed to complete cooperation, to situations involving limited communication and possibly conflicting goals. In any multiple agent scenario, it is possible that some agents may wish to cooperate, some may wish to operate independently, and others may wish to act in opposition.

Since coordination may be manifest in many forms, including cooperation, negotiation, competition, and any other form of behavior short of complete indifference and isolation between agents, we adopt a very broad definition: *Any decision by an agent that uses information concerning the existence, decisions, or decision-making strategies of any other agents is a coordinated decision.* Information used for coordination may be either known *a priori* (for example, a prediction model), it may be communicated between agents directly, it may be obtained via sensory input, or it may be learned by the agent as it functions in its environment.

Of the existing approaches to coordination, perhaps the most well known is game theory, particularly the theory of n -person, nonzero sum, noncooperative games. But one of the problems with the application of game theory to coordination is the lack of a general theory of cooperative games in extensive form. The standard solutions to games in strategic form tell us very little about conditional dynamics. While game theory may ultimately

have great application to intelligent control, that likely will not happen without overcoming many difficulties in modeling communications between players and in reconciling solutions to games played in extensive form with solutions to games played in strategic form.

Much of the recent literature on coordination deals primarily with the issue of controlling the flow of information between agents for the purpose of enhancing their decision-making capabilities. For example, communication protocols are required with hierarchical based systems [1]-[3]. Communication analysis is also required with distributed artificial intelligence problems, such as negotiation processes [4]-[6], and for distributed reasoning [7]-[12]. Research is also providing a growing body of literature on related issues such as data fusion, sensor distribution, communication bandwidth tradeoffs, reliability enhancement, system complexity, and data management. A dominant issue in coordination theory, however, is the assumption of a particular type of coordination, namely, cooperation, where all agents are united and the goal is to design an architecture that permits some acceptable joint outcome to be achieved.

The type of coordinated intelligent control discussed in this article is exemplified by the producer-consumer-transporter (PCT) problem domain introduced by Durfee [3]. Fig. 1 illustrates the PCT problem domain, consisting of a set of rooms connected by multiple doors. Each room contains two producer-consumer (PC) entities, represented by the large squares, which produce/consume two types of objects, depicted as triangular and square icons. If an icon is solid, then the object has been produced and is ready for transport; if the icon is open, then the PC is ready to receive an object. The overall objective of the procedure is to

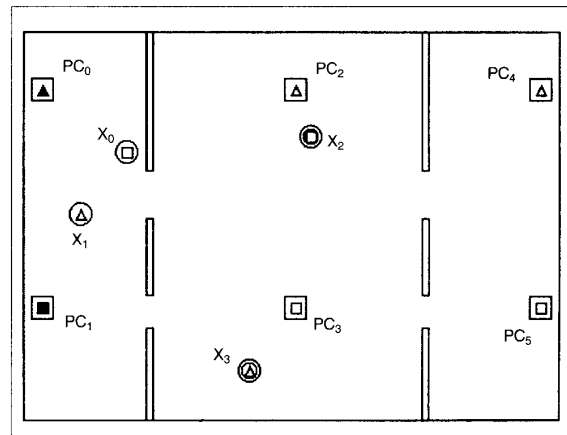


Fig. 1. Producer-consumer-transporter example.

Presented at the 1992 International Conference on Systems, Man, and Cybernetics, Chicago, IL, October 18-21, 1992. The author is with the Electrical and Computer Engineering Department, 459CB, Brigham Young University, Provo, UT 84602. Email: wynn@ee.byu.edu.

deliver icons from the left-most PC's to the intermediate room, where they are consumed and reproduced, and then from the intermediate room to the right-most room, where they are ultimately consumed. The transporters, X_0 , X_1 , X_2 , and X_3 , are represented by circles in Fig. 1. These entities are the intelligent agents of the problem, and are tasked with picking up and delivering the icons. Transporters X_0 and X_2 work as a team with X_0 servicing PC's in the left and center rooms, and X_2 servicing PC's in the center and right rooms. As time flows, the icons are produced in the left room, picked up and delivered to the center room, then picked up and delivered to the right room. Once the left room PC has had its object picked up, it is free to produce a new object. Similarly, X_1 and X_3 work as a team servicing PC's in an analogous manner. The PC's and transporters are emplaced according to some arbitrary initial conditions, and the game flows in discrete time. Although X_2 knows it is assigned to work with X_0 , it does not know initially which of the PC's in the center room is the destination of X_0 ; it must learn by observing the behavior of X_0 . X_3 has a similar situation with X_1 . We will assume the all of these agents are endowed with sensors that permit them to determine the locations of the other agents, the PC's, and of walls. Transporters are not permitted to co-occupy PC's, and they must avoid collisions with each other and with walls.

There are many possible approaches that may be used to control the behavior of the agents. If sufficient information and computational power were available, a central agent could calculate a jointly optimal course of action for all transporters, coordination would be automatic, and each agent would be given an assigned trajectory to follow. Such an approach, though optimal in the sense that it corresponds to the global minimization of some cost functional, is *nonreactive*. An alternative approach would be for the agents to possess only local control ability, and to make decisions as a function of time based only on information that is currently available. Such an approach is *reactive*, since the agents are required to respond to real-time sensor and communications information to adjust their behavior during their operation. There is no central control, there is no hierarchy, and no notion of global optimality is required. Each agent should make its decisions in accordance with some (at least local) notion of coordination, as possessed by that agent. If the agents are able to at least share some of their decision-making strategies with the other agents, then to that extent it is possible for coordination to exist.

We present here an approach for local, reactive, coordinated intelligent control. Key to the development of this approach is the concept of an *epistemic system*, together with a principle of action that may be exercised locally. Essentially, an epistemic system is composed of the agents state of knowledge, its possible actions, its system of values, and its system of beliefs. The principle of action is a mechanism to empower the agent to make decisions on the basis of these beliefs and values.

Multiple agent coordination may occur between agents by linking their epistemic systems, and by learning the epistemic systems of other agents. Two (or more) agents are linked together in this sense if at least one of them possesses sufficient information regarding other agent's epistemic system to enable itself to adjust its own epistemic system to improve its performance. Learning occurs if an agent is able to modify, through experience, its linkage such that it can improve its performance even more. We first discuss epistemic utility theory system for a single agent,

then extend it to multiple agents and describe ways of establishing linkage between separate epistemic systems.

Epistemic Systems

There are few studies in the theoretical foundations of intelligent control, and even fewer in the foundations of coordinated intelligent control. New approaches to intelligent control are beginning to emerge using expert systems, fuzzy logic, and neurocontrol, but a general framework, or set of frameworks, has not yet been established for this field in any systematic way. Since the goal of intelligent control, manifestly, is to "... autonomously do something appropriate [from a human point of view] under multiple unpredictable conditions" [13, p. 11], it may be prudent to examine anew some of the results obtained from the field of epistemology.

In the pursuit of reasonable approaches to intelligent control, we have been led to a review of some of the basic premises of decision theory. The process and methodology of human decision making and, particularly, scientific inquiry, has long been a topic of discussion and debate among philosophers. For three decades, the epistemologist Isaac Levi has actively addressed the problem of how scientific knowledge is and should be acquired. He has developed an epistemological model of the behavior of a single rational agent in the acquisition of new truth [14]-[16]. His *epistemic utility theory* addresses the broad problem of knowledge acquisition, including the goals and values of a rational agent, the agent's willingness to risk error, incomplete and contradictory evidence, the possibility that currently held beliefs are untrue, and experiment design and statistical inference. Levi's work is sufficiently broad to address real issues while providing sufficient detail and mathematical precision to be practically useful for the problems of estimation and control [17]. It parallels much of statistical decision theory in that it is based upon expected utility, but differs from conventional approaches in the definition of the utility function and in the structure of the decision rules that emerge.

For this development, we borrow liberally from [15], recognizing that our application may not be precisely what was anticipated by the founders of these concepts. Strictly speaking, epistemology deals with the issue of classifying propositions on the basis of knowledge and belief regarding their content. Epistemic utility provides a means of jointly quantifying the value of avoiding error and the value of acquiring new information, and the decision-making mechanisms that devolve from considerations of epistemic utility may be used to expand, contract, or revise one's knowledge corpus. In other words, epistemic utility is useful for making decisions regarding *what to believe*. This methodology has a natural application to the drawing of scientific inferences, making decisions under conflict or uncertainty, and in social choice.

We wish to extend the concepts of epistemic utility to decision making in the engineering context of *how to act*, and wish to apply them to machines. In so doing, may be guilty of conflating the issues of decision making and knowledge acquisition, at least when viewed from a purely philosophical point of view. Nevertheless, we find the epistemological model proposed by Levi to be a useful representation for characterizing intelligent control. In our context, epistemic utility translates into an evaluation, in terms of error avoidance and informational value, of the choices open to the machine. The mechanisms for making decisions

become principles of action, rather than simply instruments for the acquisition of new information.

Agents

An *agent* is any entity that is endowed with a body of knowledge, is capable of making decisions, and possesses goals and beliefs regarding its possible decisions. We refer to this collection of knowledge, possible decisions, goals, and beliefs as an *epistemic system* [17]. Formally, let X be an agent with epistemic system $\{K_t, U_t, G_t, B_t\}$ at time t .

K_t is X 's *knowledge corpus*, containing all logical truths, all set-theoretical truths, and all mathematical truths that are expressible in X 's language at time t , and is deductively closed.

U_t is X 's *ultimate partition*, or set of hypotheses, consisting of all seriously possible actions available at time t . We assume the cardinality of U_t is finite. Let n_t denote the number of elements in U_t , and let F_t denote the Boolean algebra of elements of U_t ; F_t contains 2^{n_t} elements. Each element of F_t is a potential answer for X .

G_t is X 's system of *informational valuation*, that reflects the informational worth or value of any action to X . This system consists of a collection of utility functions, each element of which is an equivalence class (that is, equivalent up to positive linear transformations) of utilities.

B_t is X 's a system of *truth valuation*, that reflects X 's subjective assessment of belief in each action being correct. This system consists of a *convex set*¹ of probability functions, each member of which is a serious possibility for being a valid characterization of X 's beliefs.

In the context of the PCT example, we must specify the epistemic systems of the transporter agents X_i , $i = 1, 2, 3, 4$. We will assume that the knowledge corpus for a transporter consists of the following: it knows the locations of all agents (including itself) and of the PC's. (In a more realistic scenario, this information would be obtained via noisy sensory data; we assume an idealized situation, corresponding to perfect sensors.) It knows enough mathematics to calculate distances, and it knows that it cannot collide with another transporter, and cannot pass through walls.

Each transporter's ultimate partition consists of its set of legal moves: one or two units of distance in any of the eight chess board directions (unless a move would require it to collide with something) or to remain where it is. To enable the transporters to work their way around walls, we provide them with a memory of their previous N moves, and prohibit them from moving to any of the past N locations. Each transporter's move strategy is governed by its systems of informational and credal valuation. Each transporter has three modes of action: it may move to a PC to pick up an object, it may move to a PC to deliver an object, or it may sit in an idle mode while it waits for an assignment.

Before we complete the specification of the epistemic systems for the PCT transporter agents and illustrate how they are applied to the intelligent control problem, we must first develop the

¹A set is convex if all the linear combinations with positive weights that sum to unity of elements of the set are also elements of the set. Such linear combinations are called *convex combinations*.

theory of epistemic utility and present Levi's rule of decision making.

Epistemic Utility

Under circumstances where the probability of error is so large that errors are frequent, or when actions vary in importance, or when assessments of truth are ambiguous, it is not at all clear that the selection of a single best decision, made according to some narrow optimality criterion, will lead to acceptable results. An agent may be reluctant to decide against an action that it considers important, and may be indifferent to some other action, regardless of its truth value. To deal with situations such as this, Levi proposes that the proximate aim of decision making is to *avoid error* [15]. But if decisions are to be made, then some risk of error must be worth taking. We may express this conflict as a convex combination of two utility functions: a) the utility of avoiding error, and b) the utility of obtaining new information. Let us define the utility function $\tau: F_t \times U_t \rightarrow \{0, 1\}$ as follows: for any $g \in F_t$ and $h \in U_t$, let

$$\tau_t(g; h) = \begin{cases} 1, & h \in g \\ 0, & h \notin g \end{cases}$$

τ represents the *utility of avoiding error*. That is, it is the utility of choosing g in the interest of avoiding error, if h is true.

To account for the importance of errors, we may apportion a unit of *informational value* among the elements of F_t . The potential answers may be partially ordered with respect to informational value as follows. Let $g_1 \in F_t$ and $g_2 \in F_t$ be potential answers (that is, action sets). Then accepting g_1 involves rejecting $\sim g_1$, then complement of g_1 ; similarly for g_2 . If $\sim g_1 \subset \sim g_2$ but $\sim g_1 \neq \sim g_2$, then g_1 is *less valuable informationally* than g_2 , that is, accepting g_2 rejects more elements of U_t than does accepting g_1 .

X may represent the informational value of potential answers by assigning to elements of U_t nonnegative real values such that their sum is unity. The informational value of a potential answer is the sum of the values assigned each element of U_t that is rejected according to the potential answer; that is, if we enumerate

$$U_t = \{h_{1t}, h_{2t}, \dots, h_{n_t t}\},$$

and let $m_t: F_t \rightarrow [0, 1]$ denote the value X assigns to $h \in U_t$, then

$$\sum_{j=1}^{n_t} m_t(h_{jt}) = 1.$$

For any set $g \subset U_t$, we define

$$m_t(g) = \sum_{h_{jt} \in g} m_t(h_{jt}).$$

Then $m_t(g)$ is X 's informational value of rejecting g , that is, of adding $\cup_{h_{jt} \in g} h_{jt}$ to K_t .

The function $m_t(\cdot)$ thus defined is a probability measure over F_t . This probability is termed an *information-determining* probability. It is intended to regulate the evaluation of information regardless of its truth-value — it is a measure of the demand for information. Thus, if X considers rejecting g , it is certain that it will obtain informational value worth $m_t(g)$. It does not know whether g is true or false. The utility of accepting g in the interest of acquiring new information regardless of its truth-value is, then, $c_t(g) = 1 - m_t(g)$. We may view $m_t(\cdot)$ as reflecting the goals and values of X at time t .

Given the utility function, $\tau(g; h)$, of avoiding error and the utility function, $c_t(g)$, of acquiring information, we may address the conflict that exists between these two goals by defining an *epistemic utility function* for acquiring error-free knowledge as a convex combination $u(g; h) = \alpha_t \tau(g; h) + (1 - \alpha_t) c_t(g)$. The quantity α_t represents the relative importance that X attaches to avoiding error versus acquiring new information. We must restrict

$$\frac{1}{2} \leq \alpha_t \leq 1$$

to ensure that no erroneous answer is preferred to any correct answer. This structure contrasts with the standard Bayesian formulation for the multiple hypothesis problem, where a cost C_{kj} is assessed if the k th hypothesis is chosen when the j th hypothesis is true. Thus, the standard Bayes cost structure accounts for only the cost of error, whereas the dual nature of epistemic utility permits both the cost of error and the informational value of the decisions to be specified independently. This utility function reduces to the familiar uniform cost criterion when $\alpha_t = 1$.

Since all utility functions that are related by a positive linear transformation are equivalent, we may simplify this utility function by setting

$$u^a(g; h) = \frac{1}{\alpha_t} u(g; h) - \frac{1 - \alpha_t}{\alpha_t}.$$

The resulting utility function for accepting g in the interest of both avoiding error and acquiring new knowledge is

$$u^a(g; h) = [1 - b_t m_t(g)] \tau(g; h) - b_t m_t(g) [1 - \tau(g; h)],$$

where $b_t = (1 - \alpha_t) / \alpha_t$. The system of informational valuation G_t is defined by a collection of epistemic utility functions of the form u^a that is closed under convex combinations. That is, if $u_i^a \in G_t$ and $u_j^a \in G_t$, then $\phi u_i^a + (1 - \phi) u_j^a \in G_t$ for all $\phi \in [0, 1]$.

Credal Probability and Expected Utility

Independently from establishing a utility to characterize the informational value of accepting or rejecting a given action, the agent must also establish a measure of belief for all actions that are available. We shall use probability for this measure. *Credal* probability is probability formed on the basis of subjective judgment, represents the likelihood that a hypothesis is true, and is independent of any informational value or demand that might be associated with the hypothesis. Throughout this article we shall treat the terms *credal* and *belief* as synonyms.

For a given ultimate partition U_t , let $q_t: F_t \rightarrow [0, 1]$ be a probability measure such that $q_t(g)$ denotes X 's credal probability assignment to any element $g \in F_t$. If g bears positive credal probability, then X evaluates the truth of g as a serious possibility. The system of truth valuation, B_t , is defined by a collection of credal probability functions of the form q_t that is closed under convex combinations. That is, if $q_t^1 \in B_t$ and $q_t^2 \in B_t$, then $\phi q_t^1 + (1 - \phi) q_t^2 \in B_t$ for all $\phi \in [0, 1]$.

For $g \in F_t$, the expectation of $u_a(g; h)$ may be calculated as $E_{q_t} u_a(g; h) = q_t(g) - b_t m_t(g)$. A well-accepted and rational strategy for an agent to adopt is to make decisions in a way that tends to maximize its expected utility. Thus, for any element $h_{it} \in U_t$, the strategy of accepting h_{it} bears an expected utility of $E_{q_t} u_a(h_{it}; h) = q_t(h_{it}) - b_t m_t(h_{it})$. The quantity b_t is X 's *index of boldness* at time t . Since $\frac{1}{2} \leq \alpha_t \leq 1$, we have $0 \leq b_t \leq 1$.

If $b_t = 1$, the relative importance of avoiding error is minimum. The closer b_t is to unity, the less caution X exercises that error will be introduced into its corpus; that is, it is maximally bold. The closer b_t is to zero, the lower the risk of error it will accept before being willing to expand its corpus. Thus, the expected utility represents a tradeoff between the desire to add new knowledge to the corpus and the desire to avoid error. The choice of b_t establishes the agent's threshold at which the demand for knowledge renders the risk of error worthwhile at time t .

X may adopt any hypothesis in F_t , the Boolean algebra generated by the elements of the ultimate partition U_t . This structure expands X 's hypotheses beyond the possibilities available with conventional decision theory; it is not constrained to select only the elementary events $h_{it} \in U_t$, but may choose any subset of them. The action set that maximizes expected utility of acceptance is the set $g^* = \{h_{it} \in U_t: q_t(h_{it}) - b_t m_t(h_{it}) \geq 0\}$. We can see that the set g^* can never be empty, for if it were, that would imply that

$$\sum_{i=1}^{n_t} [q_t(h_{it}) - b_t m_t(h_{it})] = 1 - b_t < 0,$$

which is impossible. Thus, the criterion of maximizing expected utility is equivalent to rejecting all action sets h_{it} such that $q_t(h_{it}) < b_t m_t(h_{it})$. If there is only one element $h_{i^*t} \in U_t$ such that $q_t(h_{i^*t}) \geq b_t m_t(h_{i^*t})$, then there is a unique maximizing (with respect to expected utility) action. If g^* contains more than one element, then none of them may be rejected on the basis of expected utility. The essence of this decision-making strategy is that, rather than selecting the one "best" action, we reject only the "bad" actions, leaving, as potential answers, all of the "good" actions for further consideration — that is, we place a higher priority on avoiding error rather than seeking the one hypothesis with smallest probability of error, which would be represented to the agent as the truth. This decision philosophy may be summarized as follows:

Levi's Rule of Epistemic Utility [15, page 53]: Given a corpus K_t , a finite ultimate partition U_t , an information-determining probability function m_t defined over the Boolean algebra of elements of U_t , an expectation-determining probability function

q_i defined over the same algebra, and an index of boldness b_i , the agent should reject all and only those elements $h_{ii} \in U_i$ satisfying $q_i(h_{ii}) < b_i m_i(h_{ii})$.

If more than one action survives this test, expected utility has failed to decide between them, and we must go beyond considerations of expected utility to resolve the question. Recall that the proximate aim of the decision strategy is to acquire new knowledge while striving to avoid error; this was the motivation for the choice of utility function, but once we have taken these principles as far as possible, a tie-breaking mechanism must be defined if it is necessary to settle on a unique decision. One conservative approach is to choose the hypothesis that maximizes the minimum utility — the maximin rule. We may apply this rule as a tie-breaking mechanism by choosing the hypothesis that maximizes the minimum utility over the set of unrejected hypotheses. That is, we extremize only over those hypotheses that have survived the tests based upon expected utility. This approach is quite different from regular maximin rules where the extremizing operations take place over the entire ultimate partition of feasible hypotheses.

For the PCT example, we interpret the informational value probability function m as representing the transporter's preference for a given move, independent of its belief of correctness. We assume that each transporter agent has a speed control throttle and, though it may accelerate to the higher speed, it is reluctant to do so because of greatly increased fuel consumption. Consequently, the agent would prefer to move at the slow speed unless there were a compelling reason to accelerate (for example, if time became critical or if it were necessary to avoid a hazard). Independently of its beliefs regarding which move is correct, however, the transporter would have no reason to prefer any move at a given speed to any other move. Assuming all possible moves for an agent are legal in the sense that they do not result in collisions, we may construct the m function to have radial symmetry, with higher values of rejection for the higher speeds. We may assign preferences as follows: Let the fuel conservational value of rejecting the hypothesis "do not move" be denoted by $m = \beta_0/n$, let the conservational value of rejecting the hypothesis "move at slow speed" be denoted by $m = \beta_1/n$, and let the conservational value of rejecting the hypothesis "move at high speed" be denoted by $m = \beta_2/n$. Since conserving fuel is preferred to expending it, the agent would assign a higher conservational value to rejecting high speed moves, which implies that $\beta_2 \geq \beta_1 \geq \beta_0$. For this example, there are seventeen possible moves for the agent: eight at high-speed, eight at low speed, and one at zero speed. To ensure that the resulting m function is a probability, we require that $\beta_0 + 8\beta_1 + 8\beta_2 = 17$. Thus, the agent gains β_2/n worth of conservational value for each high-speed move that is rejected, β_1/n worth of value for each slow-speed move that is rejected, and β_0/n worth of value for rejecting the hypothesis of not moving. In the interest of simplicity for our example, the set G_i contains only this one m function.

To determine its credal probability, each transporter computes, for each legal move, the Euclidean distance to the PC it is seeking. (We assume, in a realistic case, that this distance would be obtained from sensor data, but for our purposes we will suppose that the sensor error is extremely small.) It then subtracts this distance from the maximum over all such distances, and normalizes these quantities to unity. Thus, if we let π_i denote the

position of PC_i , the target PC, in Cartesian coordinates, and let $\{\xi_{ij}\}$ denote the set of possible moves for Transporter X_j , then the credal probability for X_j is

$$q_j(\xi_{ik}) = \frac{\max_l \{\rho_{il}\} - \rho_{ik}}{n_j \sum_{n=1} (\max_l \{\rho_{nl}\} - \rho_{in})}$$

where $\rho_{ik} = \|\xi_{ik} - \pi_i\|$. The resulting function is a probability measure defined over the ultimate partition, and assigns higher credal probability to moves toward PC_i than to moves away from it. As with the informational value determining probability, we will restrict the set G_i to contain only one credal probability.

We finish the specification of the epistemic system for each transporter agent X_i , $i = 1, 2, 3, 4$, by setting the index of boldness, b_i , to unity for each agent, thus ensuring a maximally bold strategy (in the sense that the decision rule will eliminate as many hypotheses as possible). Ties are broken according to the maximin principal applied to the unrejected hypotheses. Endowed with its epistemic system, each agent is thus empowered to function in the context of its beliefs and values within the constraints of its available moves and knowledge corpus. As play evolves in discrete time, each agent will collect and process its information through its epistemic system, make its decision, and implement its control.

Each agent is governed solely by its own epistemic system; all agents are completely autonomous. They make their decisions based upon the local conditions of the environment, and do not attempt to anticipate future behavior for themselves or for other agents. In this sense, the agents appear to be endowed with little more than "instinctual" capabilities. In fact, none of the agents is endowed with specific abilities other than what is implied through their individual epistemic systems. Clearly, much more sophisticated memory and prediction capabilities could be built into the epistemic system to make them more intelligent, but we are attempting to make this example as simple as possible without being trivial.

Coordinated Intelligent Control

For the PCT example, coordination occurs as follows: each of the transporter agents must be able to switch to the correct action mode (pick up, delivery, or wait), also, it must not collide with any other agent. In addition, Agents X_2 and X_3 must learn which PC is their pick-up point. Determination of the correct action mode may be achieved through communication with the relevant PC's, and possible collisions may be detected via local sensory output. The learning phase of coordination, however, is slightly more involved. In an ideal situation, each agent would communicate with the other agents, and would therefore know the precise assignments of its teammate. Specifically, suppose Transporter X_0 is assigned to pick up square icons at PC_1 , and that X_2 is assigned to deliver triangular icons to PC_4 . To be successful, both agents must agree on the identity of the intermediate agent, either PC_2 or PC_3 , that will be used for the relay point. Before such an agreement is established, agents X_0 and X_2 will each be in a state of tension, with credal systems

$$B_0 = \{q_{X_0}^{\alpha_0} = \alpha_0 q_{X_0|X_0} \uparrow PC_2 + (1 - \alpha_0) q_{X_0|X_0} \uparrow PC_3\}$$

and

$$B_2 = q_{X_2^2} = \alpha_2 q_{X_2|X_2 \downarrow PC_2} + (1 - \alpha_2) q_{X_2|X_2 \downarrow PC_3},$$

for $\alpha_i \in [0, 1]$; $i = 1, 2$, where $q_{X_i|X_i \uparrow PC_j}$ denotes the conditional probability mass function of X_i given that it is assigned to deliver objects to PC_j , and $q_{X_i|X_i \downarrow PC_j}$ denotes the conditional probability mass function of X_i given that it is assigned to pick up objects at PC_j . Ideal coordination between X_0 and X_2 occurs if both agents agree on their strategies; for example, if X_0 and X_2 agree to use PC_2 as their relay point. This situation would correspond to X_0 selecting $\alpha_0 = 1$ and communicating this information to X_2 , resulting in X_2 setting $\alpha_2 = 1$. A similar arrangement must be made between Transporters X_1 and X_3 relative to the origin, PC_0 , the destination, PC_5 , and the relay point, either PC_2 or PC_3 . If X_1 and X_3 are able to communicate with each other and with X_0 and X_2 , then ideal coordination will be completed if X_1 and X_3 agree to use PC_3 as their relay point.

Suppose, however, that explicit communication between transporter agents not possible, and that, consequently, X_0 is not able to communicate its value of α_0 to X_2 , who, therefore, is not able to narrow its choices for α_2 . Without additional information, X_2 would be unable to resolve this conflict, and would likely not be successful in its mission to transport objects to its assigned destination. We now suppose, however, that X_2 is able to observe the behavior of X_0 , and construct a real-time estimate of α_0 according to the following estimation rule:

$$\hat{\alpha}_0^t = \frac{t-1}{t} \hat{\alpha}_0^{t-1} + \frac{1}{t} N_t,$$

where

$$N_t = \begin{cases} 1 & \text{if } X_0 \text{ moves toward } PC_2 \\ 0 & \text{if } X_0 \text{ moves away from } PC_2 \end{cases}$$

is a discrete-time point process that monitors the motion of X_0 relative to PC_2 and PC_3 . Thus, X_2 may observe the motion of X_0 and infer the likelihood that X_0 is delivering to PC_2 or PC_3 . Once the object has been delivered (say, at time t_d , however, the learning phase is over, and X_2 must commit to one of the options, even though the iteration given above may not have converged to either $\hat{\alpha}_0 = 1$ or $\hat{\alpha}_0 = 0$. Thus, a final decision must be made before X_2 's learning is complete. This decision can also be made according to Levi's rule of epistemic utility by choosing the decision according to a likelihood ratio test consisting of the ratio of the credal and informational value probabilities associated with the decision to have X_2 choose between PC_2 and PC_3 . We will use $\alpha_0(t_d)$ as the credal probability for this test and, under the assumption that X_2 has no valuational preferences for PC_2 over PC_3 , will assign uniform informational value to these two hypotheses. Thus, X_2 will make its final assignment according to the rule

$$\hat{\alpha}_2 = \begin{cases} 1 & \text{if } \hat{\alpha}_0(t_d) \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Figs. 2-4 illustrate snapshots of computer animations of the reactive system introduced in Fig. 1. In Fig. 2, X_0 and X_1 have picked up their objects and are moving to the second room to deliver them to their destinations. Although X_2 knows that its teammate is X_0 , X_2 does not know which PC is the destination of X_0 ; consequently it observes the motion of X_0 and decides which of the two PC's in the room is its destination. X_3 undertakes a similar calculation. In Fig. 3, X_1 has successfully delivered its object, which X_3 is moving to retrieve. Similarly, X_0 is preparing to deliver its object, which will be retrieved by X_2 . Once X_1 delivers its object, it returns to retrieve a new one at its source. In Fig. 4, X_0 is carrying its object to the intermediate PC, X_1 has just delivered its object to its intermediate PC, X_2 is in the midst of transporting its object to its termination point, and X_3 is delivering its object to its termination point. Once X_2 and X_3 have completed their tasks, they will return to the intermediate PC to retrieve the next object. For this scenario, we assume that, once the agents have learned which intermediate point they are to service, it may be remembered for the duration of the scenario.

The PCT example provides a simple model for coordination between multiple agents. In this section, we present a more complex model for inter-agent coordination. Suppose each of n agents, X_i , $i = 1, \dots, X_N$, is endowed with its own epistemic system

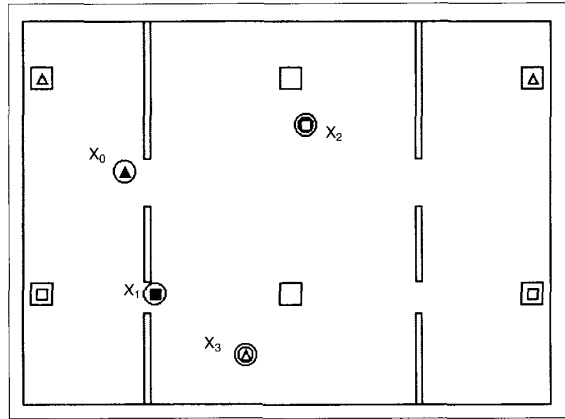


Fig. 2. PCT example: beginning of learning phase.

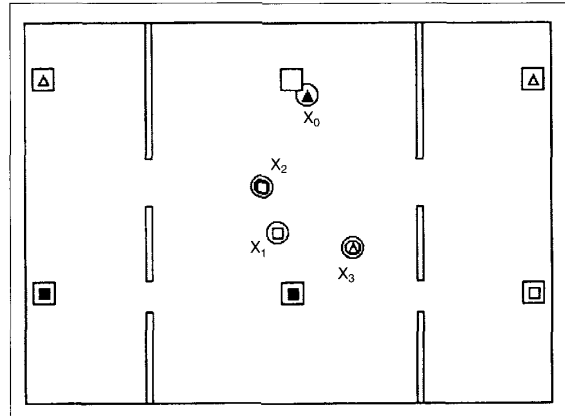


Fig. 3. PCT example: completion of learning phase.

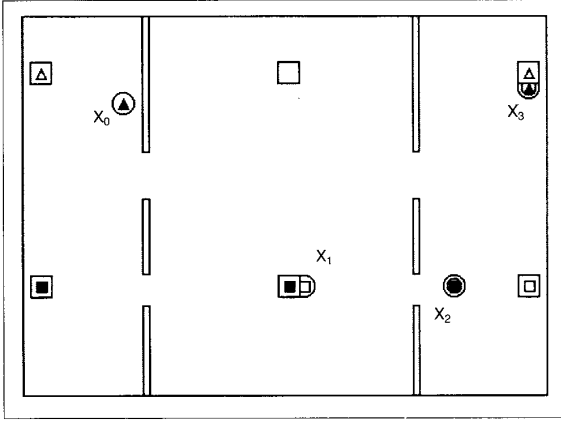


Fig. 4. PCT example: successful delivery.

quadruple, $\{K_{it}, U_{it}, G_{it}, B_{it}\}$, where $U_{it} = \{h_{it1}, \dots, h_{itn_i}\}$ denotes the ultimate partition for agent X_i at time t and F_{it} the corresponding Boolean algebra. Let $m_{it} \in G_{it}$, $m_{it}: F_{it} \rightarrow [0, 1]$ and $q_{it} \in B_{it}$, $q_{it}: F_{it} \rightarrow [0, 1]$ denote the corresponding information valuation and credal probabilities. We permit the dimensionality and the content of the ultimate partition to change at each time increment, and require only that the number of elements of U_{it} be finite. Each agent may possess a different ultimate partition, that is, we allow $U_{it} \neq U_{jt}$, $i \neq j$. Let

$$\underline{m}_{it} = [m_{it}(h_{it1}), m_{it}(h_{it2}), \dots, m_{it}(h_{itn_i})]^T$$

and

$$\underline{q}_{it} = [q_{it}(h_{it1}), q_{it}(h_{it2}), \dots, q_{it}(h_{itn_i})]^T.$$

These vectors are informational value and credal probabilities, respectively, of the ultimate partitions of all X_i , $i = 1, \dots, N$.

In general, if agent X_i has some knowledge concerning any or all elements of X_j 's epistemic quadruple, $j \neq i$, it may modify its epistemic system in at least one of two ways: a) update its q_i function (that is, change its beliefs), and b) modify its m_i function (that is, change its goals) to account for the epistemic systems of its cognate agents. Belief updating is a standard procedure in decision making. Typically, q_i will be derived from *a priori* information, sensory observations, and data fusion from multiple agents (see, for example, [18]). But updating its beliefs is not the only option open to an agent in the framework provided by epistemic utility theory. The agent also has the capability of modifying its informational value as a function of the values and beliefs of its cognate agents. This is a capability that is not typically stressed with standard maximum-expected-utility based decision-making schema. Yet there is no logical reason why an agent should be required to keep its value structure stationary while its belief structure is free to adapt. We are, therefore, motivated to develop a simple model for characterizing how an agent's epistemic system could be modified by knowledge (or estimates) of its cognate agents' epistemic systems.

To be consistent with Levi's development, we will adopt the policy that an agent's informational valuation system should be specified independently of its beliefs and vice versa. But there is no such constraint concerning the joint probabilities of multiple agents. In other words, we are not justified in assuming that the joint belief systems of X_i and X_j ($i \neq j$) factor into the product $q_{it}(\cdot)q_{jt}(\cdot)$, or that the joint belief for X_i and informational values for X_j factor into the product $q_{it}(\cdot)m_{jt}(\cdot)$, etc. If dependencies exist, then all agents could take advantage of that information by means of a multivariate version of Levi's rule of epistemic utility, namely, each agent would reject all and only those elements

$$(h_{1it_1}, \dots, h_{Nt_{i_N}}) \in U_{1t} \times \dots \times U_{Nt}$$

$$\text{such that } Q(h_{1it_1}, \dots, h_{Nt_{i_N}}) < b_t M(h_{1it_1}, \dots, h_{Nt_{i_N}}),$$

where Q and m are the joint credal and informational valuation probabilities. This formulation, while it appears to take into account all information and therefore may lay a claim to optimality, has some drawbacks. For example, it assumes that all agents actually *intend* to coordinate their actions just because they have the ability to do so, and it assumes that the multivariate distributions are known and are accessible to all agents. Perhaps the most important problem with this formulation is that, being a global approach, it may not be amenable to localized control, which is an important consideration for multiple agents. Thus, while it may be useful to think of coordination in terms of implementing a joint epistemic utility rule, it may not be practical to do so. Consequently, we seek a more flexible and perhaps more robust and useful alternative.

Joint distributions, while important for theoretical analysis, are not as useful in practice as are low-order marginal probabilities defined over small clusters of propositions [18]. This observation has also found great utility in stochastic automata theory (e.g., see [19], [20]), which is related to the development we present here. Our investigation leads us to the realization that explicit mechanisms for coordination can be introduced through conditional probabilities.

When considering how multiple agents may coordinate their efforts, it is reasonable to suppose that some of the agents would form coalitions, some would operate independently either by choice or because of insufficient information, and some would choose to be in opposition to others. A natural means of implementing these types of behavior is through the specification of marginal and conditional probabilities. We may obtain expressions for the coordinated informational and credal probabilities of each agent by defining them as convex combinations of transformed versions of the uncoordinated informational and credal probabilities. The resulting coordinated values for X_i are

$$\underline{m}_{it}^* = \sum_{j=1}^N \alpha_{ijr} A_{ijr} \underline{m}_{jt} + \sum_{j=1}^N \beta_{ijr} B_{ijr} \underline{q}_{jt}$$

and

$$\underline{q}_{it}^* = \sum_{j=1}^N \gamma_{ijt} C_{ijt} m_{jt} + \sum_{j=1}^N \delta_{ijt} D_{ijt} q_{jt}$$

α_{ijt} , β_{ijt} , γ_{ijt} , and δ_{ijt} are *coordination coefficients* to couple X_j to X_i . These coefficients, lying in the unit interval, represent weighting factors for, respectively, information-to-information coordination, credal-to-information coordination, information-to-credal coordination, and credal-to-credal coordination. A_{ijt} , B_{ijt} , C_{ijt} , and D_{ijt} are corresponding *coordination transition probability matrices* and characterize the mapping from X_j 's informational value and credal probabilities to those of X_i . To ensure that m_{it}^* and q_{it}^* are probability measures, we require that

$$\sum_{j=1}^N \alpha_{ijt} + \sum_{j=1}^N \beta_{ijt} = 1,$$

and

$$\sum_{j=1}^N \gamma_{ijt} + \sum_{j=1}^N \delta_{ijt} = 1.$$

Consistency requires that $A_{iit} = D_{iit} = I$. Also, since an agent's informational value and credal values are independent of each other, we should insist that $\beta_{iit} = \gamma_{iit} = 0$. The coordination coefficients and matrices are specified by each agent as it views its universe. These parameters may be obtained by direct communication (e.g., X_i broadcasts the message that it plans to take a specific action), they may be specified by X_i *a priori*, or X_i may compute (estimate) them conditioned upon observations taken at time t . If X_i is either indifferent to or unaware of the epistemic state of X_j , it should set the corresponding coordination coefficients to zero. If X_i chooses to make a completely uncoordinated decision, it should set $\alpha_{iit} = \delta_{iit} = 1$. The overall structure provided by this model permits decentralized decision making and graceful degradation of performance as agents' abilities to communicate decrease.

Mathematical Framework

Epistemic utility theory provides a mathematical framework in which to embed the decision-making strategies of intelligent agents. Some of the features of this approach are as follows.

- No explicit hierarchy is required. It is not necessary to specify relative authority, priority, or any other type of ranking between the agents. If a natural hierarchy happens to exist, it may be embedded into the joint epistemic system via the joint informational or belief valuations.
- The agents may be heterogeneous. They do not need to share the same decision space, the same capabilities, or the same knowledge. There is no requirement that they possess similar communication capabilities, similar sensory capabilities, or similar desires for coordination.
- The agents may be cooperative, adversative, or indifferent. Scenarios simultaneously involving cooperative agents, competitive agents, and coalitions of both types of agents may be characterized via the joint epistemic system. Even if an agent

possesses the capability of coordinating with other agents, it is not compelled to do so.

- Explicit communication between agents is not required for coordination to occur. Although some form of information sharing between agents is necessary for coordination, this sharing may take several forms: direct communication, prediction modeling, sensory observations, and learning. All of these forms of information sharing can be modeled through the epistemic system structure.

- A facility for learned coordination may be implemented within the framework provided by epistemic systems. The learning occurs when an agent is able to observe the behavior of other agents and thereby estimate some or all of the probabilities that make up its epistemic system.

References

- [1] F. Wang and G. Saridis, "A coordination theory for intelligent machines," *Automatica*, vol. 26, no. 5, pp. 833–844, 1990.
- [2] P.J. Gmytrasiewicz, E.H. Durfee, and D.K. Wehe, "The utility of communication in coordinating intelligent agents," in *Proc. Ninth Natl. Conf. Art. Intell.*, 1991, pp. 116–172.
- [3] E.H. Durfee and T.A. Montgomery, "Coordination as distributed search in a hierarchical behavior space," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 6, pp. 1363–1378, 1991.
- [4] G. Zlotkin and J.S. Rosenschein, "Negotiation and conflict resolution in non-cooperative domains," in *Proc. Eighth Natl. Conf. Art. Intell.*, 1990, pp. 100–105.
- [5] G. Zlotkin and J.S. Rosenschein, "Cooperation and conflict resolution via negotiation among autonomous agents in noncooperative domains," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 6, pp. 1317–1324, 1991.
- [6] S. Kraus and J. Wilkenfeld, "The function of time in cooperative negotiations," in *Proc. Ninth Natl. Conf. Art. Intell.*, 1991, pp. 179–184.
- [7] N. Carver, Z. Cvetanovic, and V. Lesser, "Sophisticated cooperation in FA/C distributed problem solving systems," in *Proc. Ninth Natl. Conf. Art. Intell.*, 1991, pp. 191–198.
- [8] V.R. Lesser, "A retrospective view of FA/C distributed problem solving," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 6, pp. 1347–1362, 1991.
- [9] S.E. Conry, D.J. MacIntosh, and R.A. Meyer, "DARES: A distributed automated reasoning system," in *Proc. Eighth Natl. Conf. Art. Intell.*, 1990, pp. 78–85.
- [10] S.E. Conry, K. Kuwabara, V.R. Lesser, and R.A. Meyer, "Multistage negotiation for distributed constraint satisfaction," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 6, pp. 1462–1477, 1991.
- [11] M. Klein, "Supporting conflict resolution in cooperative design systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 6, pp. 1379–1390, 1991.
- [12] W. Zhang, S. Chen, and R. S. King, "A cognitive-map-based approach to the coordination of distributed cooperative agents," *IEEE Trans. Syst., Man Cybern.* vol. 22, no. 1, pp. 103–114, 1992.
- [13] A. Meystel, *Autonomous Mobile Robots*. Singapore: World Scientific, 1991.
- [14] I. Levi, *Gambling with Truth*. Cambridge, MA: M.I.T. Press, 1967.
- [15] I. Levi, *The Enterprise of Knowledge*. Cambridge, MA: M.I.T. Press, 1980.
- [16] I. Levi, *Decisions and Revisions*. London, U.K.: Cambridge Univ. Press, 1984.

[17] W.C. Stirling and D.R. Morrell, "Convex Bayes decision theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 1, pp. 173-183, Jan./Feb. 1991.

[18] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann, 1988.

[19] K.S. Fu, "Stochastic automata as models of learning systems," in *Adaptive, Learning and Pattern Recognition Systems: Theory and Applications*, J.M. Mendel and K.S. Fu, Eds. New York: Academic, 1970, pp. 393-430.

[20] K.S. Narendra, Ed., *Adaptive and Learning Systems*. New York: Plenum, 1985.



Wynn C. Stirling received the B.A. degree (magna cum laude) in mathematics and the M.S. degree in electrical engineering from the University of Utah, Salt Lake City, UT, in 1969 and 1971, respectively. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983. From 1972 to 1975 he was with Rockwell International Corporation, Anaheim, CA, and from 1975 to 1984 he was employed by ESL, Inc., Sunnyvale, CA. Since 1984, he has been with the Department of Electrical and Computer Engineering, Brigham Young University, where he is an Associate Professor. His current research interests include decision and estimation theory, information theory, and stochastic processes. He is a member of Phi Beta Kappa and Tau Beta Pi.

Sampled Data

Answers to Control Systems Quiz

(Quiz appeared in August issue of the Magazine)

- | | |
|-------------------------------------|--|
| <u>B</u> pole-zero | A. Venus fly trap |
| <u>N</u> satellite attitude control | B. temperature for freezing in Warsaw |
| <u>F</u> root locus | C. taking a shower |
| <u>K</u> state space | D. brash, but sometimes works |
| <u>A</u> plant | E. Pope John Paul II |
| <u>D</u> direct method | F. occurs every 13 years |
| <u>J</u> unstable | G. Bode and Nyquist |
| <u>I</u> canonical | H. to be paid to the IRS |
| <u>M</u> polar pot | I. large calibre gunonical |
| <u>G</u> complex zero pair | J. matched to four unshairs |
| <u>O</u> critically damped | K. the Washington Monument |
| <u>H</u> gain margin | L. builds brick houses |
| <u>E</u> real pole | M. penguins' demise |
| <u>P</u> steady state | N. former Soviet Union |
| <u>C</u> overdamped | O. leaving the bar at 2 a.m. |
| <u>L</u> Mason | P. far from New Madrid and San Andreas |

Contributed by Louis Wozniak, University of Illinois.

Aircraft Control?

"We hope that Professor Langley will not put his substantial greatness as a scientist in further peril by continuing to waste his time and the money involved, in further airship experiments. Life is too short, and he is capable of services to humanity incomparably greater than can be expected to result from trying to fly ... For students and investigators of the Langley type there are more useful employments."

New York Times,
December 10, 1903
editorial page