Submitted to the Second International Conference on
Autonomous Agents and Multi-Agent Systems (AAMAS03)

A Preliminary Taxonomy of Multi-Agent Interactions

H. Van Dyke Parunak Sven Brueckner Mitch Fleischer James Odell

Altarum
3520 Green Ct Ste 300
Ann Arbor, MI 48105
1-734-302-4684 1-734-302-4683 1-734-302-4654
{van.parunak, sven.brueckner, mitch.fleischer}@altarum.org

James Odell Associates
3646 Huron River Drive
Ann Arbor, MI 48103
1-734-994-0833
email@jamesodell.com

ABSTRACT

Discussions of agent interactions frequently characterize behavior as “coherent,” “collaborative,” “cooperative,” “competitive,” or “coordinated.” We propose a series of formal distinctions among these terms and several others. We argue that all of these are specializations of the more foundational category of “correlation,” which can be measured by the joint information of a system. We also propose “congruence” as a category orthogonal to the others, reflecting the degree to which correlation and its specializations satisfy user requirements. Then we explore the degree to which lack of correlation can arise purposefully, and show the need to use formal stochasticity in cases where such lack of correlation is truly necessary (such as in stochastic search).

Keywords

Coordination, correlation, competition, contention, cooperation, congruence, communication, command, constraint, construction, conversation, stigmergy, agent interaction

1. INTRODUCTION

Agents often do things together. Researchers who work with multi-agent systems need to be able to discuss this joint action in a disciplined way. Yet the current vocabulary is curiously undisciplined. It is fashionable in some circles to emphasize cooperation as the dominant theme (e.g., in the title of [23] or the subtitle of [46]), but often the system designer does not care whether the agents cooperate or contend, so long as their behavior is coordinated in a certain way. In fact, market mechanisms achieve system-level coordination largely through mechanisms in which individual agents are at least competitive and sometimes contentious. Furthermore, both contention and cooperation presume a cognitive model that some architectures do not satisfy. Other terms for agent interaction include “coherence” and “collaboration” Taking advantage of the pervasive use of the Latin *co-* and *con-* in the English lexicon, we refer to them collectively (including nominal, verbal, and adjectival forms) as “Co-X,” and from this point capitalize them.

A sample of papers in the field (based on [6, 18] and papers and posters from ICMAS’95-00, Agents’97, 98, 00, and 01, and AAMAS’02) shows that from 1981 through 1993, “Cooperate”[1] accounted for eight of the twelve “Co-X” terms in titles, “Coordinate” for three, and “Coherent” for the last. “Collaborate” was introduced in 1994, and “Compete” in 1995. For 1994 through 2002, 120 Co-X terms were used, with “Coordinate” at 41%, followed by “Cooperate” at 38%, “Collaborate” at 16%, and “Compete” and “Cohere” at 2.5% each. Clearly, the community is growing dissatisfied with a simplistic characterization of agent interaction as “Cooperation.” However, when one looks beyond the titles, there is little agreement on how the various members of Co-X differ. Formal analyses of teams and social behavior (e.g., [11, 12, 13, 43]) helpfully refine concepts such as common knowledge, joint intentions, commitments, obligations, and social norms. These and related concepts certainly play a role in achieving the behavior that the Co-X vocabulary describes, but the usage of the Co-X terms themselves remains intuitive and sometimes inconsistent.

Our taxonomy embraces these and other terms and accommodates a variety of agent architectures. Such a taxonomy can enable researchers to describe more precisely how their agents interact. It may also help researchers communicate across different subfields, where different conventions may prevail. We select the particular terms we discuss based either on their common use in the literature or on our subjective assessment that they capture an aspect of agent interaction that is not otherwise covered. Under our definitions, the terms are neither a mutually exclusive spanning set such that every agent-based system belongs to exactly one term in the set (“categories”) nor an orthogonal set each of whose terms can be applied to all agent-based systems (“perspectives”). A formal taxonomy requires a complete structure of both categories and perspectives [31], but at this point we claim only, in the words of one reader of an earlier draft, “a nice start.”

Each section of this paper expounds and illustrates a part of this set. Section 2 defines the most general term, “Correlation,” in terms of a formal statistical metric over the population. Correlation makes no assumptions about either the internal structure of the agents or the relative centralization or decentralization of their behavior. Section 3 defines “Coordination” as Correlation that results from information flows from one agent to another. When these flows result from the intentions of the individual agents, we speak of “Cooperation” and “Competition,” discussed in Section 4, along with “Collaboration,” which is the intersection of Cooperation and “Conversation” (a specialization of Coordination).

AAMAS’03, July 14-18, Melbourne, Australia.

---

[1] Including the noun “Cooperation” and the adjective “Cooperative.” Throughout the paper, when we refer to one grammatical form of a given word, we intend the reader to apply our observations to the others as well.

Section 5 discusses the "Congruence" of group behavior with system-level intentions. Section 6 returns to the fundamental notion of Correlation and examines ways in which it can and cannot be avoided. Section 7 offers a summary.

2. CORRELATION: BEHAVIORAL JOINT INFORMATION

The most generic way to describe what agents do together is in terms of their joint information, otherwise known as their correlation entropy, joint entropy, or mutual entropy [2]. This quantity can be determined empirically, without access either to the internal structure of the agents or to the broader system within which they are embedded. We describe a set of agents with positive joint information as "Correlated." In some cases, such empirical observations may be sufficient to impute cognition to agents whose internal structure is unknown [8, 21, 34].

Joint information can be computed relative to many different aspects of the agents. Because we are interested in agent behaviors, we compute it over agent actions. At each time step, each agent (indexed by i) has access to any one of n_i actions $\{a_{i1}, a_{i2}, \ldots, a_{in_i}\}$. Any of our definitions could be reposed in terms of states rather than actions without affecting the point, but we prefer to focus on the agent's actions because they, unlike its state, are externally visible. Let p_{ij} be the probability that agent i executes action a_{ij}. This quantity can be estimated by maintaining a time series of the agent's last k actions (say, $k = 100$), counting the actions of each type, and dividing by k. One measure of the agent's behavior over time is its behavioral (Shannon) entropy, defined in the standard way as $H(a_i) = -\sum_{j=1}^{n_i} p_{ij} \log_2 p_{ij}$. This definition makes no assumptions about the independence of the elements in the set that generates the p_{ij}. It is simply an empirical characterization of their relative prevalence. (If k is small compared with the lifetime of the agent, one can index the p_{ij} and thus H with time t, referencing the window of length k centered at t.) Similarly, we can characterize the entropy of the overall system in terms of the various combinations of actions of the individual agents. For simplicity, we restrict our discussion to two agents, but the concepts can readily be extended to any number. The maximum total number of system actions is $n_1 * n_2$, p is now indexed over these joint actions, and the system entropy is $H(a_1, a_2) = -\sum_{j=1}^{n_1 * n_2} p_j \log_2 p_j$ (again, indexable by t).

System entropy is subadditive, $H(a_1,a_2) \leq H(a_1) + H(a_2)$. Equality obtains when the behaviors of the individual agents are statistically independent. When they are dependent, system entropy is strictly less than the sum of the individual agent entropies, and the difference $I(a_1{:}a_2) \equiv H(a_1) + H(a_2) - H(a_1,a_2)$ is the correlation, mutual, or joint entropy, sometimes called the joint information. We prefer the term "joint information," to avoid the connotations of disorder implicit in "entropy." In fact, the behaviors of agents in a system with high joint information are statistically more Correlated with one another, and in that sense more orderly, than in a system with low joint information.

Agents are Correlated when their actions are statistically dependent on those of other agents. It does not matter at this level whether this Correlation results from information flows among the agents or between them and a central controller, or whether its roots are cognitive or subcognitive. In the most general sense, Correlation will manifest itself in an increase of the system's joint information, and in fact we propose using this quantity as a measure of system Correlation. (Thus, when we speak of one system's having a higher Correlation than another, we mean that the joint information of the first is greater than that of the second.) One benefit of this perspective is that it permits us to distinguish the *fact* of Correlation from the *mechanisms* used to achieve it.

For example, suppose that each of our two agents needs access to a widget to perform its duties. Suppose further that there are two widgets available, and (to simplify the computations) that each agent accesses each widget half of the time. Then the probability that agent a_1 is accessing widget 1 is $p_{1,1} = 0.5$, as is the probability that agent a_1 is accessing widget 2, the probability that agent a_2 is accessing widget 1, and the probability that agent a_2 is accessing widget 2. We further assume that a widget works better when only one agent is using it at a time. Each individual agent entropy is $-2*0.5 \log_2(0.5) = 1.0$. At the system level, there are four possible joint actions: both agents accessing widget 1, both accessing widget 2, a_1 accessing widget 1 and a_2 widget 2, and *vice versa*. If the agents do not Coordinate their activities, then each of these four possibilities is equally likely, with probability 0.25, and the system entropy is $-4*0.25 \log_2(0.25) = 2$. This value is equal to the sum of the individual agent entropies, so the joint information is 0. Now assume the agents' behaviors are Correlated (by whatever means) so that they avoid the joint actions in which they both choose the same widget. In this case there are only two system actions, each with probability 0.5, and the system entropy is 1, less than the sum of the agent entropies. The difference, 1, is the joint information between the agents.

We illustrate the application of joint information to a model of resource allocation, the minority game, described in more detail in [36]. Briefly, at each time step each of the N agents in the population (where N is odd) seeks to allocate itself to one of two resources. Each agent receives a point for each turn that it finds itself on the less-occupied (minority) resource, and the system goal is to maximize the total points awarded across the entire population of agents (or, equivalently, to minimize the variance in the population of either resource over successive runs). The information available to the agents is a time series identifying which resource was in the minority at each past cycle. In each turn of the game, the agents consult the last m entries in this time series, and use this to choose the resource they will access on the next time step. The quantity $z = 2^m/N$ reflects the normalized size of the strategy space available to the agents, and turns out to be a critical parameter in the dynamics of the game.

Figure 1 plots the normalized variance σ^2/N as a function of z. Since low variance reflects high system-level payoff, desirable behavior is located at the minimum of this curve, where $z \approx 0.34$. The dashed line shows the behavior that would result if all agents made random choices. The minimum is a spin-glass phase transition, discussed in more detail in [26, 38]. As z decreases below this point, the system performance quickly degenerates until the agents are doing worse than if they made random choices. In this region, study of the time series of minority groups [25] shows "herding" behavior. Relatively few distinct strategies are available for small m, with high probability that agents' decisions will overlap. Above the phase transition, the agents do better than random, but as z increases, performance approaches the random limit as an asymptote.

Consider a population of 61 agents. The entropy of each agent is computed on the basis of its probability of choosing resource 0 or 1 at each step, while the entropy of the system is computed on the basis of the probability of a specific vector of 61 individual agent choices at a time step. Thus the system as a whole has a state space of $2^{61} \approx 10^{18}$. Reasonable experimental runs with this system are on the order of 10^4 to 10^6 steps, so it would be irresponsible to estimate probabilities over this state space for the whole system from our experimental data. As an alternative, we focus on subsystems of six agents each. Such a subsystem has a state space of 2^6, over which we can reliably estimate probabilities with experiments of $10^4 \approx 2^{13}$ steps. Thus each run of 5001 steps lets us look at ten subsystems of six agents each, and we conduct thirty runs in all.

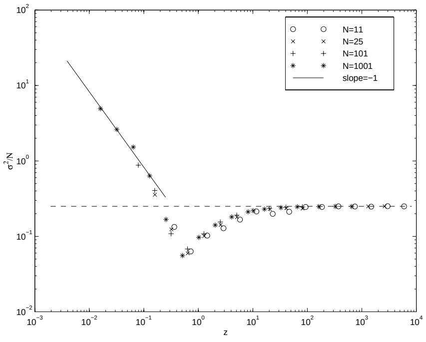

Figure 1: Performance vs. Size of Strategy Space in the Minority Game.—σ^2/N measures system inefficiency (lower is better). $z \equiv 2^m/N$ is the size of the strategy space accessible to the agents. The dashed line indicates random choice by the agents.

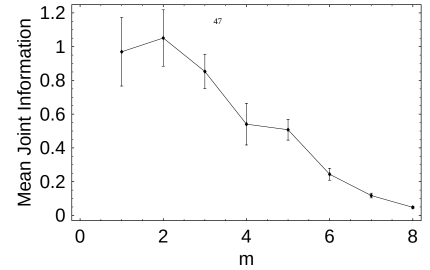

Figure 2: Joint Information in the Minority Game.—Error bars mark one standard deviation. The phase transition in Figure 1 corresponds to m = 4.37.

Figure 2 shows the Correlation, as measured by the joint information. This figure has three important features, corresponding to the three regions of Figure 1.

1. The Correlation is highest for low *m*, consistent with the analysis in [25] showing herding behavior in this region. Within the low region, the mean value appears to increase from *m* = 1 to *m* = 2 before declining for *m* > 2, but given the size of the standard deviation in region, the most that can be said is that the Correlation is comparable for these two values. There are few distinguishable strategies available to the agents for low *m*, resulting in higher Correlation among their behaviors.
2. High *m* is associated with low Correlation, corresponding to the region where agent decisions approach the random limit. Thus the general shape of the curve is logistic: low slope for low and high *m*, and steep slope in between.
3. There is a deviation from this general shape between *m* = 4 and *m* = 5, the region that corresponds to the phase transition (which would be at *m* = 4.37, though our experiments do not sample this point).

Thus the joint information reflects the overall structure of Figure 1. However, the two measure different things. Figure 1 measures an estimate of the overall system performance, while Figure 2 measures Correlation. These two measures are not statistically correlated, a point that we discuss further in Section 5.

3. COORDINATION: COMMUNICATION

Perhaps the most commonly used term for agent interaction is "Coordination." For example, it is used in the ACM computing classification system under I.2.11, "Distributed Artificial Intelligence" [1] (paired with "Coherence," which we discuss under "Congruence" below). The difference between "Correlation" and "Coordination" is that "Correlation" describes simply the fact of statistical non-independence among agent behaviors, while "Coordination" implies a causal process. Correlation can emerge among randomly generated numbers as a statistical fluke, but when it arises from a causal process, that process involves communication, that is, information flow between an individual agent and its environment. The environment, in turn, is everything that lies outside the individual agent's boundary. The options for this flow stem from the contents of this environment, which include both other agents and environmental state variables. (A side effect of this definition is to require redefinition of either "Coordination" or "communication" in paper titles of the form, "Coordination without communication" [3, 15, 16, 40].)

The environment includes other agents, not only software agents but also human stakeholders, system designers, and conventional computer systems. The relationship between a given pair of such agents at any point in time will be one of two types, depending on the state of the agents and the rules of the system (expressed, for example, in the agents' roles [35] and the protocols in which they participate). When the agents can say "No" to one another within the rules of the system, we say that they are "peer agents." When one of them (say agent A) can say "No" to the other (B), but B cannot say "No" to A, we call A the "distinguished agent" and B the "subordinate." The relationship between two agents may be fairly fixed (for example, the relationship between a human programmer and her software agent). Or it may vary over time (as when peer agents negotiate a work plan that calls for one of them to supervise the other, resulting in a distinguished-subordinate relationship during execution). These concepts can be developed more formally through dependency and autonomy theory [10, 30].

The environment has state variables that can be sensed by the agents, and may support its own processes that couple its state variables and cause them to change over time. We distinguish two categories of environmental variables. The values of *endogenous* variables change over time depending on the actions taken by the peer agents. The values of *exogenous* variables, such as sunspot frequency, change over time independent of the actions of the peer agents, but may be viewed as resulting from actions of the distinguished agent, and are often scripted by the system designer.

In such an environment, Correlation mechanisms can be either centralized or decentralized. The fact that Correlation describes the results of both mechanisms enables us to frame disciplined comparisons between centralized and decentralized solutions to the same problem. In addition, the information flows involved may be either direct from agent to agent (ignoring those aspects of the environment that make up the communication system) or indirect (mediated through explicitly modeled environmental state variables). Table 1 reflects these categories.

Centralized mechanisms for Correlation all involve communication between the distinguished agent and its subordinates. This flow may take place directly (when the distinguished agent Constructs or Commands the subordinates) or indirectly (when

Table 1: Categories of Communication

		Topology of Inter-Agent Relationships	
		Centralized (between Distinguished and Subordinate agents)	**Decentralized** (among Peer agents)
Information Flow	**Direct** (messages between agents)	***Construction*** (Build-Time) ***Command*** (Run-time)	***Conversation***
	Indirect (non-message interaction)	***Constraint***	***Stigmergy***[2] (generic) ***Competition*** (limited resources)

the distinguished agent Constrains the subordinates by manipulating exogenous environmental variables visible to the subordinates). In Correlation through Command, used commonly in robot soccer, holonic manufacturing, and some simulation applications, agents behave much like objects, executing methods invoked by incoming messages. The focal point algorithm advocated by [15] and the common utility functions implicit in [16] both rely on Construction (common programming). In indirect centralized mechanisms, subordinates jointly sense changes in a shared exogenous environmental variable. The variable's dynamics are independent of agent actions, so it cannot move information between subordinates. But it may serve as a synchronizing signal that Correlates the agents' actions. The experimenter who configures targets and obstacles in an experimental testbed is Constraining the subordinates, supporting Correlation through indirect centralized action.

Decentralized mechanisms for Correlation all involve communication among peers. Most negotiation research focuses on direct peer-to-peer information flows ("Conversation"). Indirect decentralized flows occur when peers make and sense changes to endogenous environmental variables. This class of Coordination is called "stigmergy," [17], from the Greek words stigma "sign" and ergos "work": the work performed by agents in the environment guides their later actions. Such techniques are common in biological distributed decentralized systems such as insect colonies [32]. A common form of stigmergy is resource Competition, which occurs when agents seek access to limited resources. For example, if one agent consumes part of a shared resource, other agents accessing that resource will observe its reduced availability, and may modify their behavior accordingly. Even less directly, if one agent increases its use of resource A, thereby increasing its maintenance requirements, the loading on maintenance resource B may increase, decreasing its availability to other agents who would like to access B directly. In the latter case, environmental processes contribute to the dynamics of the state variables involved. (We reserve "Competition" for resource Competition as a subset of Stigmergy. For the more generic opposite of "Cooperation," we prefer "Contention," discussed below.)

We became aware of the active role that the environment plays in negotiation when experimenting with an instance of the contract net for manufacturing control [29]. After carefully proving that our protocol was deadlock-free, we ran it on a physical control system, and it promptly deadlocked. The negotiation in question concerned the movement of a physical part from one workstation to another. Our analysis of the protocol neglected the movement of the physical part itself. This movement conveyed information between the two workstations, and thus between the software agents that represented them. It represented an undocumented extension of our protocol, one that invalidated our proof and caused the system to deadlock. The arrival of the part at the receiving workstation gave that workstation information about the state of the system that it would not otherwise have had, namely, that the part had been delivered.

Traditionally, the study of negotiation focuses on Coordination by means of information flow directly from one agent to another. The mantra of situated robotics that "the world is its own best model" [7] suggests that the problem domain may deserve a more prominent role in the process. There are several motives for understanding the role of the environment in Coordination, and learning to exploit it where possible.

- It supports open, heterogeneous societies of agents. The environment is by definition accessible to the agents that are negotiating about it. Any agent that wishes to deal with the domain must be able to sense and manipulate it. Thus the physics of the environment define common standards for agent interaction, in contrast with the more arbitrary standards programmers can impose on direct agent-to-agent communication.
- It integrates and reflects the state and dynamics of the overall problem-solving process at a global level that is only imperfectly visible in the internal models maintained by any of the agents. In particular, it captures high-order interaction effects that may escape the notice of any individual agent or *a priori* model maintained by an individual agent. For instance, assume agents A, B, C, and D are all interested in resource ℵ, but A and B know only of each other, as do C and D. The load on resource ℵ integrates information about the demands of all the agents that would otherwise not be available to them.
- It embeds domain constraints (e.g., resource limitations) directly in the reasoning process, without the need to identify and model them in advance.

A stock market illustrates the importance of information flows mediated by endogenous environmental variables. It affects both stock traders and business executives, in different ways. Traders (at least those who obey SEC regulations) do not communicate directly to determine which shares each will buy and sell. But when a trader offers for sale a share in one company, the offer tends to depress that company's share price, making the company more attractive to potential buyers. Thus information flows between traders through the stock market without Conversation. In contrast, business executives rely extensively on Conversation in reaching contracts with their customers and suppliers. However, they must also pay attention to indirect information flows, including those through the same stock market. For example, if a supplier's stock price drops precipitously, the supplier may not be able to raise needed capital, and in spite of its explicit promises in a negotiation, it may not be able to fulfill its obligations.

The Minority Game is an excellent example of Stigmergy, and of Competition in particular, and we discuss its implication for indirect Coordination further in [36].

These mechanisms reflect Coordination mechanisms recognized by sociologists in organizational design. One prominent discussion [27] distinguishes five such mechanisms:

1. Mutual adjustment, informal communication among workers, corresponds to the Direct Decentralized quadrant of Table 1, which we call "Conversation."

---

[2] Unlike other terms in Co-X, "Stigmergy" does not begin with "Co," but the term is too well established in the research community to replace it with an alliterative alternative.

2. Direct supervision is our “Command,” which represents the real-time portion of our “Direct Centralized” quadrant.
3. Standardization of work processes (e.g., setting up a work station on an assembly line) is adjusting the environment to Constrain agents to behave in a certain way, and thus corresponds to our “Indirect Centralized” quadrant.
4. Standardization of outputs insures that intermediate outputs from one worker can be input to the next. This mechanism enables “stigmergy,” our “Indirect Decentralized” quadrant.
5. Standardization of skills and knowledge trains workers to behave in Coordinated ways. This is our “Construction,” in the “Direct Centralized” quadrant of the table.

4. COOPERATION[3] AND CONTENTION: INTENT

Determining *Correlation* across a population of agents is an empirical process that requires no knowledge about either the internal structure of the agents or their outward organization. The focus on communication processes emphasized by *Coordination* requires that we investigate inter-agent organizational issues, but still leaves the internal logic of the agents undefined. To determine whether agents are Cooperating or Contending, we must inquire into their intentions. For example, traders in a commodity market exhibit a high degree of Correlation in their actions, resulting from information flows among them (thus Coordination). But are they Cooperating or Contending? Two traders both bidding for the same commodity might be Contending (each seeking to wrest control from the other), Cooperating (pumping the price up to increase the value of their current holdings), or simply Competing (in the sense defined in Section 3). The difference can only be resolved by determining the intent of the participating agents. Thus our definition supports that of [24].

We do not have space to work out a full theory of Cooperation and Contention, but suggest that a necessary condition for Cooperation is the existence across the Cooperating agents of joint intentions ([11]). Similarly, contention suggests an intention on the part of one agent to frustrate the intentions of another. To our knowledge, this notion of “antagonistic intent” has not been formalized, but could be along the same lines as joint intention. We do not require intention for Competition, respecting the common use of the term for discussing agents seeking common limited resources without harboring malice toward one another.

The definition of Cooperation and Contention as Correlation driven by agent intent has two important implications.

First, imputing Cooperation and Contention to agents also requires imputing cognition to them. This requirement is most directly met for Cognitive agents, as in the SOAR [28] or BDI [37] architectures, which seek to imitate the representations and processes of human cognition. It is much less direct for Behavioristic agents. Such agents, inspired by work in artificial life [33], are “black boxes,” defined only by their outward behavior, and their internal programming makes no claims to imitate the detailed functioning of cognition. In fact, it is possible to impute cognition to such agents in a disciplined way [34], but doing so would require observing other behaviors beyond the specific actions to be classified as Cooperation or Contention, in order to deduce their intentions toward one another.

Second, neither Cooperation nor Contention requires direct decentralized communication (“Conversation” in Table 1), but might result from centralized design-time information flows or environmentally mediated interactions. We distinguish the special case when agents both Converse and Cooperate as “Collaboration,” and use “Coalition” to describe the resulting state of affairs.

5. CONGRUENCE AND COHERENCE: USEFULNESS

None of the modes of interaction discussed thus far is necessarily desirable. Consider the simple problem of Correlated access by two agents to two widgets considered in Section 2. There, we saw that if the agents avoid choosing the same widget, they are Correlated, achieving a joint information of 1. Of course, the agents will be just as Correlated if they always choose the same widget, but in this case productivity would be lower in the Correlated system than in the random one. Similar examples can be constructed to illustrate that increased Coordination, Cooperation, and Competition do not always result in more productive systems.

The crucial insight here is that systems can have goals associated with them at two levels: the system, and the individual agents. (Cognitive agents reason explicitly about these goals, while in behavioristic agents they are imposed by the agent designer, but they are still agent-level goals.) Categories such as Contention and Cooperation take into account individual agent goals, but not system goals. We propose “Congruence” to characterize the degree to which the pattern of agent interactions (at any level from Correlation through Contention and Cooperation) satisfies (“is Congruent with”) system-level goals. “System-level” is critical. For example, in an e-commerce system, each individual agent may have a different user with different goals (e.g., increased market share vs. short-term profit). Congruence deals, not with the conformity of individual agents to the goals of their respective users, but with the conformity of the system as a whole to its system-level goals (e.g., bounded transaction times, information availability, and transaction security).

The relation among the agents themselves that yields Congruence is “Coherence” (Figure 3). This latter term appears (without definition) in the ACM Computing Classification [1]. (Durfee et al. [14] define “Coherent” as “well-coordinated.”)

System-level goals can arise in two different ways. In an engineered system, they are defined by the system’s creators [41]. We term these “top-down goals.” In a cognitive multi-agent system, they can also emerge from agent interactions [42], whether through democratic processes or by the imposition on other agents of the individual goals of an agent that has gained a controlling position in the society. These goals are thus “emergent goals.” Several points need to be made.

1. Congruence does not presuppose either peer-to-peer information flows or individual agent intent. It may exist with any form of Correlation.
2. One can conceive of systems that do not have system goals, and for which Congruence cannot be defined. Like intentions, emergent goals are most naturally associated with cognitive agents. Congruence can be postulated of behavioristic agents in three ways. First, their creator may define their system-level goal. Second, if they exist in a larger system that includes both non-cognitive and cognitive agents, the cognitive agents may

---

[3] Axtell [4] notes that game theoreticians would reverse our definitions of “coordination” and “cooperation.” We believe our definitions are more in line with the usage in the MAS community. The exact words used are much less important than precision in distinguishing the processes involved.

define emergent system goals for the entire system, and do their best to impose them on the non-cognitive portion of the system. Judging from the persistence of ants, mosquitoes, and cockroaches, such efforts may meet with varying success. Third, such goals may be imputed to them per [34].

3. A system may have conflicting goals. These may arise in populations of non-cognitive agents from inconsistency in the designer's goals, in populations of cognitive agents from tensions among different emergent processes, and in created systems of cognitive agents from a disjunction between the designer's top-down goals and goals that emerge from within the population. Congruence is defined only with respect to a specified goal or set of goals.
4. System-level goals are needed to *define* Congruence, but whether or not they *affect* it depends on whether individual agents can sense and respond to them. Emergent goals that do change the behavior of the system exemplify "downward causation" [39].

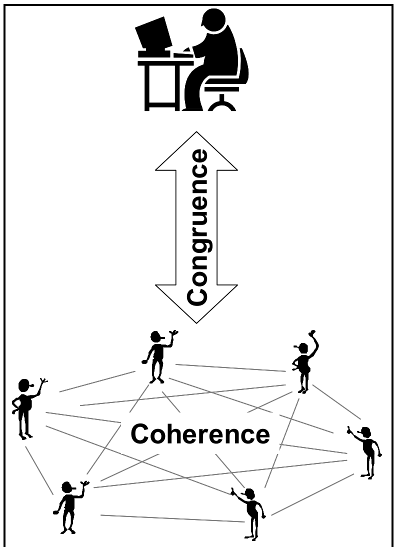

Figure 3: Congruence and Coherence.—*Congruence* expresses alignment between system behavior and system goals. *Coherence* is the relation among agents that produces Congruence.

The minority game exemplifies a top-down system goal (maximum total points awarded across the population). This goal is not downwardly causative, because the agents do not know of it or reason about it, and the system is Congruent only in the vicinity of the phase transition. Changes in Correlation (Figure 2) reflect the coarse structure of system performance (Figure 1), but are not statistically correlated with them. In particular, the highest level of Correlation (and thus Coordination) occurs for low *m*, while the system is most Congruent and the agents most Coherent for intermediate levels of *m*.

System-level goals, under rubrics such as "norms," "conventions," and "obligations," are the subject of considerable study in the MAS community ([13] and references there). Space does not permit a detailed theory, but whatever such theory one may define will set a standard against which to assess Congruence.

6. ANTICORRELATION

The various members of Co-X are all refinements of Correlation, whether defined by information flows (Coordination), individual intent (Cooperation and Contention), or system-level goals (Congruence and Coherence). The underlying assumption (already challenged by our discussion of Congruence and Coherence) is that more Correlation is a good thing. How do these forms of interaction manifest themselves in a situation in which either the peers (individually or corporately) or a distinguished controller seek to eliminate Correlation?

Such a situation might arise in at least three ways.

1. The agent system might be in Contention with an adversary that could take advantage of observed regularities in its performance. In such a situation, the system should seek to avoid regularities, and appear as though it were made up of statistically independent entities.
2. Due perhaps to similarities in their internal coding, the agents may tend to "run into" each other in the problem space, and need to spread out to do their job effectively.
3. The agent system may be using some form of weak search (such as particle swarm optimization [22] or evolutionary computation [19]) to search the space of system behaviors. Such mechanisms rely on an assumption of ergodicity: they depend on the dynamics of the system to sample the state space, and Correlation among agents implies that there are regions of the state space that are not represented in the sample.

The default way for a set of agents to anticorrelate is for each to make its decisions on the basis of random processes. A central controller could run the random process and direct the agents' actions accordingly, or each agent could run its own random process. Can they be more deliberate about it, using either peer-to-peer or master-slave information flows to guide their decisions? We can make two observations.

1. Any system in which at least two of the agents are Correlated is Correlated. To see this, let $A = \{a_i\}$ be the set of peers, $B = \{b_i\} \subset A$ the subset that is Correlated, $H(A)$ the entropy of the entire system, and $H(a_i)$ the entropy of the *i*th peer. Perfect anticorrelation requires $H(A) = \sum H(a_i)$, where the sum is over the elements of A. Each agent can contribute at most $H(a_i)$ to this sum. In particular, each element of B must make its full contribution to the sum, requiring $H(B) = \sum H(b_i)$, where the sum is now over elements of B. But B is Correlated, so by definition of Correlation, $H(B) < \sum H(b_i)$. Each b_i contributes less than $H(b_i)$ to $H(B)$ and thus to $H(A)$, so $H(A) < \sum H(a_i)$, and the entire system is Correlated.
2. If the set of anticorrelating agents has more than one member, they must use a random process in their decision-making to achieve anticorrelation. To see this, assume that the agents do not use random processes. Then their actions are a deterministic function either of a non-random central signal or of observations (direct or indirect) of one another's behavior. But then each agent's behavior is not statistically independent of the actions of the other agents, and the system will be Correlated.

Observation 1 makes it very unlikely that MAS researchers will ever deal with perfectly anticorrelated systems. Correlation wants to happen. If agents are behaving in any way other than randomly, their aggregate behavior will reflect it. To put it another way, emergent behavior is ubiquitous. This behavior may not be Congruent, but it will be Correlated. A simple example is herding behavior in financial markets. Some researchers suggest that emergent behavior is a threat to be suppressed by constraining the behavior of individual agents so that the system exhibits only a subset of its total potential behavior [9, 20, 45]. A more realistic approach is to understand the mechanisms that drive emergence so that we can harness it for productive use.

Observation 2 emphasizes the importance of stochasticity as an element of multi-agent systems. If we want agents to spread out through their joint state space, we can do no better than to have them flip coins. In the parlance of statistical mechanics, such a device provides the "symmetry breaking" that avoids undesirable Correlation. A system's level of organization is inversely proportional to its level of symmetry [5], and random variations among agents is a powerful way to introduce differences that can be amplified by agent interactions to yield self-organization. Many techniques of swarm intelligence [32] include a stochastic element. Ant path planning requires that ants not follow pheromone gradients absolutely, but that they weight a random walk based on the gradient. Swarm sorting algorithms pick up and deposit items, not deterministically, but based on the Fermi function of recently observed concentrations. The emergence of organization in *Polistes* wasps depends on a stochastic transfer of force between agents in which the stronger wasp usually, but not always, wins the face-off. Elsewhere [34] we exhibit a simple artificial agent whose performance is dramatically improved by addition of random noise.

7. CONCLUSION

Agents do things together. Clear discussions of what they do, and effective designs of how to do it, require precision in the terms we use to describe joint behavior. We suggest the following ontology.

The fundamental characteristic is *Correlation*, defined as nonzero joint information over a population of agents. Agent Correlation is a purely behavioral notion. It requires knowledge only of the observed actions of the agents. If we admit other sorts of knowledge, we can refine it in three orthogonal ways.

Coordination is Correlation with a focus on the information flow that enables it, and six different flavors can be distinguished: Conversation, Construction, Command, Constraint, Stigmergy, and Competition. The main distinctions are whether the information flow is centralized or peer-to-peer, and direct or indirect Thus Coordination implies a particular architecture between agents, but is silent about their internal processing.

Cooperation and *contention* are Correlation modulated by the intent of individual agents. Cooperation requires joint intentions, while Contention requires an intention on the part of one agent to frustrate the intentions of another. Both of these concepts require us to impute cognition to the participating agents (thus requiring special care in the case of behavioristic agents), but they are silent regarding the inter-agent architecture, and thus independent of Coordination. A system with both Conversation (direct peer-peer communications) and Cooperation (joint intent on the part of the individual agents) exhibits *Collaboration*, which results in *Coalitions* of agents.

Congruence measures the degree to which an agent system aligns with a system-level goal, which may be defined either endogenously or exogenously. It is independent of both inter-agent and intra-agent architecture. *Coherence* is the relation among agents that yields Congruence. Importantly, Congruence is not necessarily a monotonic function of Correlation. Sometimes increased Correlation (or Coordination, or Cooperation) may yield lower Congruence.

This preliminary taxonomy can and should be extended in a number of dimensions. For example, it is fruitful to consider temporal distinctions in the ways agents work together, such as synchrony vs. asynchrony [44]. More disciplined attention to these distinctions will enable researchers to communicate more precisely just what a multi-agent system can achieve, and will help users select more intelligently from among available technologies.

8. ACKNOWLEDGMENTS

This work is partly supported by DARPA ANTS and NA3TIVE under contracts F30602-99-C-0202 and N00014-02-C-0458 to Altarum. The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the US Government.

9. REFERENCES

[1] ACM. The ACM Computing Classification System [1998 Version]. 1998. HTML, http://www.acm.org/class/1998/.

[2] C. Adami. *Introduction to Artificial Life*. New York, NY, Springer Telos, 1998.

[3] R. C. Arkin. Cooperation without Communication: Multi-agent Schema-Based Robot Navigation. *Journal of Robotic Systems*, 9(3):351-364, 1992.

[4] R. Axtell. Personal communication 2002.

[5] P. Ball. *The Self-Made Tapestry: Pattern Formation in Nature*. Princeton, NJ, Princeton University Press, 1996.

[6] A. H. Bond and L. Gasser, Editors. *Readings in Distributed Artificial Intelligence*. San Mateo, CA, Morgan Kaufmann, 1988.

[7] R. A. Brooks. Intelligence Without Representation. *Artificial Intelligence*, 47:139-59, 1991.

[8] B. Browning, G. A. Kaminka, and M. M. Veloso. Principled Monitoring of Distributed Agents for Detection of Coordination Failures. In *Proceedings of Distributed Autonomous Robotic Systems (DARS-02)*, 2002.

[9] S. Bussmann. Agent-Oriented Programming of Manufacturing Control Tasks. In *Proceedings of Third International Conference on Multi-Agent Systems (ICMAS'98)*, pages 57-63, IEEE Computer Society, 1998.

[10] C. Castelfranchi. Founding Agent's 'Autonomy' on Dependence Theory. In *Proceedings of 14th European Conference on Artificial Intelligence*, pages 353-357, IOS Press, 2000.

[11] P. Cohen and H. J. Levesque. Teamwork. Technical Report Technote 504, SRI International, Menlo Park, CA, 1991.

[12] F. Dignum and B. v. Linder. Modelling social agents: Communication as actions. In M. Wooldridge, J. Muller, and N. Jennings, Editors, *Intelligent Agents III*, vol. 1193, *LNAI*, pages 205-218. Springer-Verlag, New York, NY, 1997.

[13] F. Dignum, D. Morley, E. A. Sonenberg, and L. Cavedon. Towards Socially Sophisticated BDI Agents. In *Proceedings of Fourth International Conference on MultiAgent Systems (ICMAS'2000)*, pages 111-118, IEEE Computer Society, 2000.

[14] E. H. Durfee, V. R. Lesser, and D. D. Corkill. Coherent Cooperation among Communicating Problem Solvers. *IEEE Transactions on Computers*, C-36:1275-1291, 1987.

[15] M. Fenster, S. Kraus, and J. S. Rosenschein. Coordination without Communication: Experimental Validation of Focal Point Techniques. In *Proceedings of International Conference on Multi-Agent Systems (ICMAS'95)*, pages 102-108, AAAI, 1995.

[16] M. R. Genesereth, M. Ginsburg, and J. S. Rosenschein. Cooperation without Communication. In *Proceedings of National Conference on Artificial Intelligence (AAAI'86)*, pages 51-57, AAAI, 1986.

[17] P.-P. Grassé. La Reconstruction du nid et les Coordinations Inter-Individuelles chez *Bellicositermes Natalensis et Cubitermes sp.* La théorie de la Stigmergie: Essai d'interprétation du Comportement des Termites Constructeurs. *Insectes Sociaux*, 6:41-84, 1959.

[18] M. N. Huhns and M. P. Singh, Editors. *Readings in Agents*. San Francisco, CA, Morgan Kaufmann, 1998.

[19] C. Jacob. *Illustrating Evolutionary Computation With Mathematica*. San Francisco, Morgan Kaufmann, 2001.

[20] N. R. Jennings. On Agent-Based Software Engineering. *Artificial Intelligence*, 117:277-296, 2000.

[21] G. A. Kaminka, M. Fidanboylu, A. Chang, and M. Veloso. Learning the Sequential Behavior of Teams from Observations. In *Proceedings of RoboCup Symposium*, 2002.

[22] J. Kennedy, R. C. Eberhart, and Y. Shi. *Swarm Intelligence*. San Francisco, Morgan Kaufmann, 2001.

[23] V. Lesser and D. D. Corkill. Functionally accurate, cooperative distributed systems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-11:81-96, 1981.

[24] A. Lux and D. Steiner. Understanding Cooperation: An Agent's Perspective. In *Proceedings of First International Conference on Multi-Agent Systems (ICMAS'95)*, pages 261-268, MIT and AAAI, 1995.

[25] R. Manuca, Y. Li, R. Riolo, and R. Savit. The Structure of Adaptive Competition in Minority Games. Program for the Study of Complex Systems, University of Michigan, Ann Arbor, MI, 1998. URL http://www.pscs.umich.edu/RESEARCH/pscs-tr.html.

[26] M. Marsili, D. Challet, and R. Zecchina. Exact solution of a modied El Farol's bar problem: Efficiency and the role of market impact. 1999. PDF File, http://ttt.lanl.gov/PS_cache/cond-mat/pdf/9908/9908480.pdf.

[27] H. Mintzberg. *Structure in Fives: Designing Effective Organizations*. Englewood Cliffs, NJ, Prentice-Hall, 1993.

[28] A. Newell. *Unified Theories of Cognition*. Cambridge, MA, Harvard University Press, 1990.

[29] H. V. D. Parunak. Manufacturing Experience with the Contract Net. In M. N. Huhns, Editor, *Distributed Artificial Intelligence*, pages 285-310. Pitman, London, 1987.

[30] H. V. D. Parunak. Distributed AI and Manufacturing Control: Some Issues and Insights. In Y. Demazeau and J.-P. Müller, Editors, *Decentralized AI*, pages 81-104. North-Holland, 1990.

[31] H. V. D. Parunak. Hypercubes Grow on Trees (and Other Observations from the Land of Hypersets). In *Proceedings of The Fifth ACM Conference on Hypertext*, pages 73-81, ACM, 1993.

[32] H. V. D. Parunak. 'Go to the Ant': Engineering Principles from Natural Agent Systems. *Annals of Operations Research*, 75:69-101, 1997.

[33] H. V. D. Parunak, S. Brueckner, and J. Sauter. ERIM's Approach to Fine-Grained Agents. In *Proceedings of NASA/JPL Workshop on Radical Agent Concepts (WRAC'2001)*, pages Forthcoming, 2001.

[34] H. V. D. Parunak and S. A. Brueckner. Imputing Agent Cognition from Dynamics. In *Proceedings of Autonomous Agents and Multi-Agent Systems (AAMAS 2003)*, pages (submitted), 2003.

[35] H. V. D. Parunak and J. Odell. Representing Social Structures in UML for Agent-Oriented Software Engineering. In *Proceedings of Workshop on Agent-Oriented Software Engineering*, pages 17-24, 2001.

[36] H. V. D. Parunak, R. Savit, S. A. Brueckner, and J. Sauter. Experiments in Indirect Negotiation. In *Proceedings of The AAAI Fall 2001 Symposium on Negotiation Methods for Autonomous Cooperative Systems*, 2001.

[37] A. S. Rao and M. P. Georgeff. Modeling Rational Agents within a BDI Architecture. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning (KR-91)*, pages 473-484, Morgan Kaufman, 1991.

[38] R. Savit, R. Manuca, and R. Riolo. Adaptive Competition, Market Efficiency, Phase Transitions and Spin-Glasses. PSCS-97-12-001, University of Michigan, Program for the Study of Complex Systems, Ann Arbor, MI, 1997. URL http://xxx.lanl.gov/abs/adap-org/9712006.

[39] R. K. Sawyer. Simulating Emergence and Downward Causation in Small Groups. In *Proceedings of Multi-Agent-Based Simulation (MABS'2000)*, pages 49-67, Springer, 2000.

[40] S. Sen, M. Sekaran, and J. Hale. Learning to Coordinate Without Sharing Information. In *Proceedings of National Conference on Artificial Intelligence (AAAI'94)*, pages 426-431, AAAI, 1994.

[41] Y. Shoham and M. Tennenholtz. On Social Laws for Artificial Agent Societies: Off-Line Design. *Artificial Intelligence*, 73:231-252, 1995.

[42] Y. Shoham and M. Tennenholtz. On the Emergence of Social Conventions: modeling, analysis and simulations. *Journal of Artificial Intelligence*, 94(1-2):139-166, 1997.

[43] G. Tidhar, E. A. Sonenberg, and A. S. Rao. On Team Knowledge and Common Knowledge. In *Proceedings of Third International Conference on Multi-Agent Systems*, pages 301-308, IEEE Computer Society, 1998.

[44] D. Weyns and T. Holvoet. Synchronous versus Asynchronous Collaboration in Situated Multi-Agent Systems. In *Proceedings of Autonomous Agents and Multi-Agent Systems (AAMAS 2003)*, pages (submitted), 2003.

[45] M. J. Wooldridge and N. R. Jennings. Pitfalls of Agent-Oriented Development. In *Proceedings of 2nd Int. Conf. on Autonomous Agents (Agents-98)*, pages 385-391, 1998.

[46] M. Yokoo. *Distributed Constraint Satisfaction: Foundations of Cooperation in Multi-Agent Systems*. Berlin, Springer, 2001.