08-028

Recognizing the New: A Multi-Agent Model of Analogy in Strategic Decision-Making

Giovanni M. Gavetti
Massimo Warglien

RECOGNIZING THE NEW:

A MULTI-AGENT MODEL OF ANALOGY IN STRATEGIC DECISION-MAKING *

Giovanni Gavetti
Strategy Unit
237 Morgan Hall
Harvard Business School
Boston, MA 02163
(617) 495-5378
ggavetti@hbs.edu

Massimo Warglien
Dept. of Economics and Bus. Management
Ca' Foscari University of Venice
Cannaregio 873
30121 Venice, Italy
041 2348745
warglien @unive.it

October 2007

RECOGNIZING THE NEW: A MULTI-AGENT MODEL OF ANALOGY IN STRATEGIC DECISION-MAKING

Abstract: In novel environments, strategic decision-making is often premised on analogy, and recognition lies at its heart. Recognition refers to a class of cognitive processes through which a problem is interpreted associatively in terms of something that has been experienced in the past. Despite recognition's centrality to strategic choice, we have limited knowledge of its nature and its influence on strategic decision-making in individuals, much less in the multi-agent settings in which these decisions typically occur. In this paper, we develop a model that extends neural nets techniques to capture recognition processes in groups of decision-makers. We use the model to derive some fundamental properties of collective recognition. These properties help us understand how the intensity of communication among group-members and some select structural characteristics of the group affect recognition outcomes in novel and structurally ambiguous worlds. In particular, we demonstrate that communication pressure can lead agents to converge to shared interpretations or recognitions that are new to each of them, thereby helping them recognize problems that are genuinely new. We also show that when communication is too intense, its beneficial aspects give way to the pathologies of "groupthink." We conclude by discussing how our results are relevant to strategic choice, as well as how our model complements both other theories of choice that view the role of experience as central and recent work in population ecology that emphasizes cognitive processes.

Keywords: recognition, analogy, collective decision-making, strategic choice, cognition, novel environments.

INTRODUCTION

Strategic decision-making is most salient when firms face novel environments. It is particularly during times of change, in the early phases of a new industry, or after a discontinuity of some sort that firms must discover and pursue viable strategic positions. At these times, the context of choice is typically hard to interpret: among other reasons, knowledge of cause and effect relationships is unavailable or difficult to obtain, the nature of industry participants is ambiguous, and opportunities are ill-defined. In the language of decision theory, *states of the world* and the probabilities over such states are not naturally given, and the strategic decision-maker cannot easily define them. Decision-makers are thus *structurally ignorant.* What underlies the intelligence of choice in these settings? We argue that *recognition* is essential to such choices. Recognition refers to a class of cognitive processes through which a problem or situation is interpreted associatively in terms of something that has been experienced before. In this paper, we develop a formal model that attempts to capture realistically the cognitive phenomenology of recognition for both individuals and groups. The model allows us to identify some of recognition's basic properties, and thereby to offer a set of answers to the question we posed above. Our argument follows from two premises.

First, cognitively plausible accounts of strategic decision-making in novel worlds require us to move away from the imagery of rational choice, which portrays "model-based anticipation of consequences evaluated by prior preferences" (March, 2006: 202). When agents are structurally ignorant, they can neither easily build models of hard-to-construct states of the world nor evaluate consequences based on such models (Gilboa and Schmeidler, 2000). Structural ignorance instead suggests an imagery of experience or analogy-based decision-making, in which agents transfer wisdom or solutions that have a history of favorable outcomes from contexts that they believe are similar to the situation at hand. Central to this imagery is a *logic of recognition*, which places the interpretation of the new in terms of the old at the center of decision-making.[1] The general advantage

---

[1] Although we use analogy and recognition almost interchangeably, the two concepts do not fully overlap. Analogy is the transfer of wisdom from past situations that are considered similar to the problem at hand. Thus, analogy comprises two processes: recognition

of recognition is that it does not entail structural knowledge about the decision problem's structure (Gick and

Holyoak, 1980; Thagard, 1996). Instead, it involves some mapping between elements central to the old

experience's structure and features of the new problem that the decision-maker can readily "see," and being

cognizant of such features does not require the agent to have in-depth knowledge of the problem's underlying

causal structure (March and Simon, 1993: 10-13). In the domain of strategy, an example of this form of

reasoning is the adoption of a multi-brand strategy by Lycos, an Internet portal firm. After Lycos' managers

chose to grow the firm quickly by acquiring other Internet firms, they decided Lycos should maintain multiple

brands rather than become a monolithic entity such as Yahoo!. As documented by Gavetti and Rivkin (2007),[2]

Lycos' managers arrived at this solution after an off-site meeting during which they recognized that Lycos was

operating in a setting that was similar to the media industry. As such, they viewed Lycos as a media company,

and successful media companies such as Time Warner traditionally adopt multi-brand strategies. Lycos'

managers did not attempt to evaluate the likely consequences of various strategies in the Internet portal industry.

Rather, the cognitive basis for their decision was analogy: they mapped from a "source" context of prior

experiences to the current "target" context, and recognized the target in terms of the experience they envisioned

as being most similar (Gick and Holyoak, 1980).

Second, strategic decisions are rarely the product of individual strategists. As the Lycos example suggests,

they frequently involve group effort, such as that of a top-management team or a consulting team. If analogical

forms of reasoning are salient in novel situations, and group effort is the norm, the recognition of novel worlds is

likely to occur collectively, as the result of multiple agents' interactions. Such dynamics might affect the nature

and quality of the recognition process, and therefore the nature and quality of the decisions stemming from it.

Groups of analogizers can vary on several dimensions, including the overlap across their experiences, the

---

(the interpretation of the "new" in terms of the "old," generally referred to as "similarity mapping") and transfer (the transfer and adaptation of past solutions to the new situation). In this sense, recognition is a central component of analogy. In this paper, we focus on the recognition aspect of analogy, and refer to recognition and analogy as if they were synonymous – what we mean by both terms is the process through which a new situation is interpreted, and therefore recognized, by its association with past experiences.

[2] See also "Lycos (A): The Tripod Decision," HBS case study, Gavetti, Rivkin, and Johnson, 2002.

intensity of their communication, and the structure of their interaction. Little is known, however, about how such variables affect recognition and thus analogy's quality.

Although analogy and therefore recognition at both the individual and group levels are vital to decision-making in settings that are relevant to strategists, strategy scholars have been relatively silent on the topic. Even recent work that develops the cognitive foundations of strategic choice (cfr., for instance, Simon, 1991; Ocasio, 1997; Camerer and Lovallo, 1999; Huff and Huff, 2000; Gavetti and Levinthal, 2000; Denrell, Fang, and Winter, 2003), or, more broadly, work on organizational cognition and sensemaking (cfr., for instance, Weick, 1995; Meindl, Stubbart, and Porac, 1996; Lant and Shapira, 2001) has largely neglected the role of analogy. Analogy-based forms of strategic decision-making fall in the middle ground of *semi-rationality* between the polar perspectives of deductive economic reasoning (Brandenburger and Stuart, 1996; Porter, 1996; Ghemawat, 1999) and local, problemistic search (Cyert and March, 1963; Mintzberg, 1978; Levinthal, 1997; Winter, 2000). Yet, except for Gavetti, Levinthal, and Rivkin's (2005) initial foray, little is known about how strategy-makers can navigate this middle ground, or how they can exploit the virtues of analogy and avoid its perils. In contrast, decision theorists (Gilboa and Schmeidler 2000), political scientists (Neustadt and May, 1986), and negotiation scholars (Thompson et al, 2003) have explored analogy's potential more extensively.

To sum up: analogy and its recognition logic are central to strategic choice, which typically occurs in multi-agent contexts. We have limited knowledge about individual-level recognition in strategic decision-making, and virtually no knowledge about how its properties translate to multi-agent decision-making settings. These premises represent our point of departure, and the gaps they highlight chart the territory of our effort. In this paper, we derive a cognitively plausible model of the recognition processes that underlie strategic choice in novel settings. Our model seeks cognitive realism at three levels.

First, we focus on recognition, thereby plausibly characterizing the cognitive mechanisms involved in interpreting novel realities, about which agents are structurally ignorant. More specifically, we characterize how

agents with a limited memory of past experiences recognize the current problem in terms of this memory. Our modeled agents' recognition is governed by a memory that operates in associative terms over sets of features (Anderson and Bower, 1980). That is, these agents represent the problem at hand in terms of qualitative features, and recognize it by choosing the past experience that most closely matches the features they use to represent the current problem.

Second, at a more micro level, we rely on models that mimic the neural mechanisms that are thought to underlie associative processes. Specifically, we rely on the formal apparatus of associative neural networks (Hopfield, 1982; Amit, Brumel, and Tsodyks, 1994), which is thought to capture the basic properties of the neural processes involved in associative memory (see Miyashita, 1988, and Fuster, 1995, for neurophysiological demonstrations of such properties, and Amit, Brumel, and Tsodyks, 1994, and McRae, de Sa, and Seidenburg, 1997, for attempts to reproduce experimental observations of human memory by associative neural network models).

Finally, we seek to realistically characterize decision-making settings by moving from the individual to multi-agent settings. Specifically, we use this microfoundation of analogical thinking at the individual level to explore collective processes of recognition. Building on Hutchins (1995), we construct "networks of neural networks" -- groups of communicating individual agents who recall memories and interpret their environment collectively. On this basis, we explore a few select dimensions of how groups are structured and operate, and these dimensions' implications for the intelligence of recognition.

Through this formal structure, we induce some core properties of recognition in groups of individuals. In particular, the model suggests that groups' communicative pressure is a central determinant of recognition outcomes. We demonstrate that, within certain ranges, communication pressure leads agents to converge to shared interpretations or recognitions that are new to each of them. That is, these recognitions do not correspond to any prototype already stored in individual memories. This property is obviously important when agents face

novel problems that do not match any of their previous experiences: thanks to communication, novelty can be recognized as such. Communication pressure can also lead, however, to pathological outcomes. We demonstrate how, when its intensity is too strong, classic "groupthink" properties may emerge.

The paper is structured as follows. In §2, we discuss our assumptions about recognition processes. In §3, we describe the model. §3.1 describes the model of individual-level recognition, which we extend to multi-agent settings in §3.2. In §4.1, we demonstrate, in closed form, some basic properties of the model. In §4.2, we articulate a more complex and realistic model in which agents have heterogeneous knowledge of their environment and signals from the world are noisy. Via computer simulation, we compare the performance of different communication structures and show some non-monotonic effects of communication on group performance. Finally, in §5 we discuss the model's implications and avenues for future work. In particular, we discuss complementarities and associated opportunities for opening a dialogue between our approach and other conceptions of choice and action, such as case-based decision theory (Gilboa and Schmeidler, 2001) and appropriateness-based notions of choice and routines (March, Schulz, and Zhou, 2000; March and Olsen, 1976), which emphasize the importance of experience and associative processes. We conclude by briefly discussing the potential relevance of our work to recent developments in population ecology (e.g., Hannan, Pólos, and Carroll, 2007), which have increasingly come to view cognitive processes akin to the kind that we consider here as central to how organizational forms emerge.

PREMISES, ASSUMPTIONS, AND MODELING CHOICES

The proposition "Problem B can be recognized as being similar to Problem A; thus, solution X, which was used successfully in A, can be transferred to solve problem B" represents the basic logic of analogical reasoning. We will call B the *target* problem, and A the *source* problem. The core of analogy is the process by which the target is recognized as being similar to an experience the agent stores in her memory. How do individuals map from past experiences to current ones? How do they access their memory to retrieve experiences and associated

5

lessons that can be applied to the new context? Less abstractly, how can strategic decision-makers assess the similarity between, say, an emerging (and therefore novel and structurally unknown) industry and those they have experienced, either directly or indirectly? Our model focuses on this retrieval and mapping process, which underlies recognition. It builds on two central assumptions.

First, we assume recognition is based on associative memory. Models of analogy-based decisions commonly assume some kind of exhaustive "brute force" search through individual memory. For example, case-based decision theory (Gilboa and Schmeidler, 2001) assumes that agents look through *all* cases stored in their memory, and Gavetti, Levinthal, and Rivkin's (2005) computational model adopts a similar perspective. In the spirit of cognitive realism, we take a slightly different path, and assume retrieval is driven by associative memory (Anderson and Bower, 1980). Associative memory is a "content-addressable" process that, instead of involving the serial search through all possible direct and vicarious experiences that are stored in the agent's long-term memory,[3] directly retrieves one that closely matches the target along the dimensions guiding the search process. A classic example of associative memory is the way a flavor or smell can instantly evoke entire episodes from the past: a simple flavor can immediately lead to the recognition of a situation in terms of ones previously experienced, and not all memories have to be searched for these experiences to be evoked.

Second, following a long tradition in cognitive psychology (Tversky, 1977; Rumelhart and Ortony, 1977; Gentner, 1983), we model individual representations of situations as clusters of features. Relatedly, consistent with work on categorization (Rosch and Mervis, 1975; Rosch, 1978), our model assumes that individual memory is organized in terms of prototypical situations, or experiences, that correspond to different clusters of correlated features. These features can be specific attributes of the situation, which are typically referred to as object features, or structural relationships among such attributes, which are typically referred to as structural

---

[3] Long-term memory contains both direct and vicarious experiences. Students of memory refer to direct experiences as "episodic memory," which comprises memories of specific episodes directly experienced by the agent. Yet long-term memory also comprises a semantic component. "Semantic memory" refers to "memory for facts, the vast network of conceptual information underlying our general knowledge of the world" (Sutton, 2003). For instance, to stay with the media industry example, the long-term memory of a strategy-maker might contain direct, episodic experiences of the media industry, or semantic knowledge of its central attributes, which the agent acquired vicariously.

features (see Gentner, 1983). For instance, in the media industry, typical examples of *object* features might be the nature of customers, the role of advertising, the nature of the product, the type of competition, et cetera. In contrast, a typical example of a *structural* feature might be "advertising is key because intrinsic product quality is hard to assess, and therefore customers' taste is easily shapeable." Independent of these features' natures, Lycos' managers would have stored the prototype *media industry* in their memory as a cluster of features. Although competing theories of analogical reasoning have disagreed on how much object features contribute vis-à-vis structural features when experiences stored in individual memory are accessed, there is experimental evidence that both types of features are important for accessing memory and retrieving experiences, with individuals tending to focus more on object features (Catrambone, 2002).[4]

Neural networks techniques offer a family of formal approaches to capture both associative processes and feature-based conceptions of representations and memory (Hertz, Krogh, and Palmer., 1991). These techniques are empirically robust: when used to reproduce experimental observations on human memory, they consistently offer strong explanatory power (Amit et al., 1994 and McRae et al., 1997).[5] Among the various neural network approaches, the Hopfield model (Hopfield, 1982) is the simplest and mathematically best understood model of content-addressable memory retrieval. As we explain below, Hopfield's formal architecture represents both novel situations (the target) and prototypes stored in the agent's memory (the source) as clusters or networks of correlated features. In this model, the memorized situation that is most similar to the input stimuli (the features of the target that the agent "sees") is retrieved through a process that exploits correlations among features to converge to a single memory pattern without "visiting" the agent's full memory, with a remarkable economy of

---

[4] This debate is paralleled by a related debate on the relative *effectiveness* of object attributes versus relational ones in assessing similarity between source and target domains. Some scholars (e.g. Tversky, 1977) argue that the higher the overlap among object features is, the higher the similarity between a given source and target is. That is, ceteris paribus, if situation A shares a higher number of object attributes with situation B than it does with situation C, it will be more similar to situation B than it is to situation C. Others (e.g. Gentner, 1983) argue that structural features (i.e., relationships among features) offer a more reliable basis for similarity mappings. Our position is that, whenever possible, structural features are preferable to object features in assessing similarity. Representations that focus on relationships among features are more likely to capture the true causal structure underlying a given situation, thereby offering a deeper basis for similarity mappings. At the same time, structural mapping imposes a heavier burden on the individual: it requires a deeper understanding of some features of the target's causal structure, which may be difficult to obtain in novel situations. Despite their obvious importance, our model abstracts from these prescriptive considerations.

[5] Our claim about neural networks' empirical robustness is limited to their explanatory power vis-à-vis associative memory tasks. These models have performed less effectively when used to represent other cognitive functions (Pinker and Prince, 1988).

processing. Because the Hopfield model exhibits a combination of empirical robustness, correspondence with our assumptions, and mathematical tractability, we consider it appropriate for deriving a model of the recognition processes that underlie analogical reasoning.[6]

THE MODEL

In this section, we outline a cognitively grounded model of how groups of agents collectively recognize or interpret new situations or realities. We then use this structure as the microfoundational basis for deriving our group-level model.

The basic setup: Individual level

Basic intuition. Our model of the individual builds on the Hopfield model, arguably the prototype of most associative neural networks (Hopfield, 1982; Hertz, Krogh, and Palmer, 1991). The agent in this model is given a novel input (i.e. the target problem or situation) that she needs to recognize and that she represents in terms of features that might or might not exist. For instance, in the media industry, "economies of scale in marketing" might be present, but "buyers' switching costs" might be absent. Recognition is modeled as an associative process, and thus rests on the agent's memory, which in turn is modeled as a collection of prototypical experiences (or source problems). Experiences are stored in the individual memory as a neural network, with each node of the network representing a feature, and the connections among nodes or features encapsulating the agent's experiences. Specifically, the higher the correlation between two given features across the agent's experiences is, the "heavier" the connection between such features is. That is, if features x and y tend to be jointly present or jointly absent across the agent's experiences, their connection will be heavier than it would be

---

[6] From a neurophysiological perspective, it is important to recognize that neural network models, including the Hopfield model, offer only a partial representation of the brain's complex phenomenology. Nevertheless, even in their simplest form, they seem to capture some basic properties of the brain's functioning beyond what we noted (Anderson, 1995; Hopfield, 1982; Hertz, Krogh, and Palmer, 1991; Smolensky and Legendre, 2006). In particular, they capture what is regarded as a central mechanism underlying cognition: how information is transmitted across neurons, resulting in neurons' activation or inhibition. Recently, there have been attempts to blend basic neural associative memory with high-level symbol processing to seek higher levels of cognitive realism. Some of these models (cfr. Kokinov and Petrov, 2001 and Hummel and Holyoak, 2003) are particularly interesting for their attempt to provide an explicit formal account of structural features in analogy. These attemps are promising, and merit special attention. For our purposes, however, particularly given our intent to characterize multi-agent settings, which implies an extra layer of analytical complexity, we privilege mathematical tractability and simplicity, and thus focus on the simpler Hopfield model.

if the two features did not co-occur. With this setup in mind, it might be useful to regard the nodes of the network as hypotheses about the presence or absence of corresponding features. When the agent is given a new situation to recognize, the network will be initialized to reflect the new situation (e.g., each node will reflect hypotheses about the presence or absence of features according to how the agent perceives the new reality). This event is the starting point of an updating process during which the network will modify its configuration, thereby correcting the agent's initial perception of the reality, according to the consistency between the hypotheses the network is currently considering and the agent's memory. This iterative process will continue until the network converges to one of the experiences the agent stores in her memory.

A simple visual example will help clarify the nature of the associative recall process. Consider a memory that stores two prototypical patterns, as in Figures 1a and 1b.

Each configuration can be represented as a vector of 25 "nodes" (the little squares of which patterns are made). Once the memory is stimulated by a new visual input (Figure 1c), it progressively modifies the state of its nodes until it converges after a few iterations to the stored configuration that matches most closely the new input (Figure 1e).

The model has three basic components: environmental states (i.e., the new reality to be recognized or target problem), which are coded as *situations* (i.e., configurations of features); agents' memories, which we model as limited repertories of situations stored in associative neural networks; and network dynamics, which are triggered when new environmental states that start associative recall are presented as inputs to agents. We describe them in turn.

Situations. Following the first premise we laid out above, we assume that agents ordinarily represent situations (new realities to be interpreted or experiences in the agent's memory) as networks of features. We assume that the set of such features, F= {$f_1, f_2, .. , f_n$}, is finite (n=N). Thus, each situation can be encoded by a

vector **s** of n binary state variables that take on value 1 when the feature is present and –1 when it is absent. Consequently, there are 2^N conceivable situations.

Memory. Individuals store a repertoire of situations in their limited memories (see Gilboa and Schmeidler, 2001 for a closely related assumption). We assume that the set of situations stored in an individual memory (M) is a subset of the set Z of all conceivable situations, and that the former has much lower cardinality than the latter does. Thus, M⊂Z and #(M)<<#(Z).

We model an individual memory as a neural network, which is made of nodes (i.e., the "artificial neurons") that can "fire" or become active when incoming stimuli exceed some threshold. Consistent with the interpretation we suggested above, a given node or neuron fires when the hypothesis is accepted that the feature associated with it exists. Nodes are connected by arcs (i.e., the "artificial synaptic connections") that pass stimuli from node to node. In our model, there is a node for each feature, and the network is fully connected by symmetric arcs (see Figure. 2). The network graph can be conveniently translated into a pair (**s,W**) where **s** represents the nodes' states, and **W**, the matrix of weights, represents the adjacency network of the network graph. Formally, **s** is a vector of N binary variables s_i that take on values {1,-1}, and **W** is a symmetric N×N matrix of real-valued weighted connections w_{ij}.

INSERT FIGURE 2 ABOUT HERE

Network dynamics. The neural network has dynamics represented by how, given a matrix of connection weights **W**, nodes update their state once a new situation is presented as an input to the network. When a new situation is presented, each node takes as its initial state the value consistent with the state of the corresponding situation. In other words, the set of features (as perceived by the agent) is directly translated into the nodes' activation. Then, the update process is based on a classical principle of neural network models: each node s_i of the network takes as an input the activation state of each other node j≠i, weighted by the strength of the connections from the j-th node to the i-th node. At this point, such inputs are simply summed up. If the sum of

inputs is above a given threshold, the node becomes (or stays) active. Otherwise it becomes (stays) inactive. The

update process proceeds sequentially for each node s_i. The simplest way to model this principle in a network in

which a node state of 1 stands for activation and –1 stands for inactivity is to take the sign of the aggregate input

to determine the value of the i-th node (see also Figure. 3):

$$(1) \quad s_i = \text{sgn}(\sum_j w_{ij} s_j)$$

Rule (1) has two useful properties:

1) It has been proved that it always leads the network to a configuration in which no node state is perturbed

any longer (a fixed point). Because of this property, the agent's memory can be represented as a set of situations

stored as fixed points.

2) When a new input is presented to the network, rule (1) guarantees convergence to the memorized

situation that is most similar to the perceived input in terms of features (technically speaking, to the fixed point

with the lowest Hamming distance). Thus, memory recall is a feature-matching process that associates new

situations with memorized ones according to similarity. Visually, the stored situations can be represented as

decomposing the space of conceivable situations in basins of attraction that are determined by this similarity

metric. Figure 4 (adapted from Hertz, Krogh, and Palmer, 1991) shows an idealized 2-D representation of such

decomposition in basins of attractions around the stored situations.[7]

Appendix 1 offers a simple example of the model's basic mechanics.

Storing situations. Although nodes update their state during the recall process, the matrix **W** of connection

weights is kept constant. Thus, the connection weights are the parameters of the network dynamics. This

condition implies that a situation can be memorized by appropriately tuning the **W** matrix. In the language of

---

[7] The space of features is a highly dimensional hypercube in which the corners are the binary states of feature variables. The two-dimensional Euclidean space representation in Figure 4 only hints at the "basins of attraction" imagery.

neural networks, this memorization process corresponds to the tuning of the synaptic connections' strength. There are two ways to store situations in agents' memory. One is to allow them to learn the weights experientially (for instance, through learning procedures such as the Hebbian rule, which provides an effective way to learn the appropriate weights (Hertz, Krogh, and Palmer, 1991)). Alternatively, we can directly "engineer" the storage of situations as fixed points in agents' memory. Because our focus is not on how agents learn to memorize situations, but rather on how a given memory is used to interpret new situations, we adopt the latter approach.[8]

From the individual to the group

Elements of the model. Building on the model for individual agents, we characterize the group as a set of agents with a communication structure among them. Specifically, agents in a group share a common coding of the environment (i.e., they have the same number of nodes, corresponding to the same set of features), but may differ in terms of their memories' content (e.g. the repertoire of stored situations). Further, we assume that agents: a) have no conflicting interests; b) can differ in terms of the environmental features they pay attention to; c) can occupy different positions in the communication structure.

We model communication among agents through the architecture introduced by Hutchins (1995) and further developed by Marchiori and Warglien (2005). In this architecture, agents communicate by sending signals to each other about their current recognition of the environment (as represented by the current configuration of their memory network). Signals focus on specific features (nodes). For instance, if agent 1 and 2 communicate, and agent 1 believes f_i is present, she transmits a signal to agent 2 about the presence of f_i. This

---

[8] One way to do so (Hertz, Krogh, and Palmer, 1991) involves defining **W** as the sum of the "outer products" of each stored situation vector $\mathbf{s}^k$ with itself. In formal notation:

$$\mathbf{W} = \sum_i s^i \circ s^i$$

where $\circ$ is the outer product of two vectors

message adds to the input received by agent 2 on f_i. Figure 5 shows a simplified, two-agent communication structure (dotted lines), in which agents communicate over two features.

This structure allows us to control three basic parameters. First, communication can be more or less *intense*, which we capture by modulating the weight of the signal transmitted among agents. Thus, high intensity corresponds to a high impact of the messages to the receiving agent. Second, communication can be more or less *dense*, as dictated by which agent talks to which other agent within a group. Third, communication among agents can differ in *scope.* That is, a pair of agents can communicate over a more or less extensive set of features.

The last two parameters, density and scope of communication, allow us to model alternative group structures that correspond to different communication patterns. For each group structure, we can then tune intensity of communication. For instance, we can capture a "*full network structure*" of communication by having each agent communicate with each other. Alternatively, a "*star*" structure can be obtained by having each agent communicate to only a "central" agent. In turn, these structures can take on two forms: communication is *complete* if agents communicate on all features; it is *specialized* if they communicate on a subset of features, perhaps reflecting cognitive division of labor, with agents' areas of expertise being complementary. Because we focus on group structures in the simulation analysis, we provide more details in section 4.2.

Formal Specification. The group-level model could be described as a "network of networks." In fact, the whole group is itself a larger associative memory net. Consider n agents with m feature-nodes each. Each agent k is associated with a vector of states $\mathbf{s}^k$ of length m. Appending such vectors to each other will generate a vector $\mathbf{s}$ of length m*n, which represents all nodes in the group. Each node of $\mathbf{s}$ will be connected to both "within-agent" nodes and, via communication, other-agents' nodes. We will keep the symbol w_{ij} (which, when

necessary, is supplemented by a superscript for each agent k) to represent within-agent connections, and we will

use the symbol γ to indicate the intensity of a generic between-agent connection. For simplicity, and without loss

of generality, we will assume that the value of γ (and therefore the intensity of communication) is the same for

all between-agent connections. In concrete terms, the parameter γ indicates how influential an agent *k*'s

interpretation of a given feature is over another agent's interpretation of the same feature, with high levels of γ

meaning high influence, and low levels denoting low influence. Given our assumption that it is the same for all

between-agent connections, γ reflects a group-level property of communication – the level of influence that

members of the group have on each other. By tuning γ, we can therefore obtain groups characterized by various

degrees of "internal influence," which correspond to various degrees of communication intensity (e.g., *low*

influence/weak communication vs. *high influence/intense communication* groups). γ can be interpreted as a

proxy summarizing the potentially infinite and diverse factors that determine the influence of communication

among group members.

The matrix **W** of all such connections is a symmetric square matrix that has the (shaded) blocks constituted

by each agent's internal connections **W**^k (i.e., individual memories) on its main diagonal, and the blocks

representing communication among agents outside the main diagonal (see Figure 6).

<div align="center">INSERT FIGURE 6 ABOUT HERE</div>

Once communication is introduced, the update rule (1) for a single feature-node of a single agent becomes:

$$(2) \quad s_i^k(t+1) = \text{sgn}\left\{ \sum_j w_{ij}^k s_j^k(t) + \gamma \sum_{\substack{p=1 \\ p \neq k}}^{n} s_i^p(t) \right\}$$

ANALYSIS

Collective Recognition: Three fundamental questions (and associated properties)

Our model is designed to explore groups of decision makers' collective recognition of new realities, particularly regarding how the environment and group characteristics (such as intensity of communication, structure of interaction, and level of cognitive heterogeneity among group members) affect recognition.

By "collective recognition," we mean that the group converges to a stable state in which no agent has reason to modify individually her current interpretation of the environment. We do not require all agents to share the same interpretation -- they might agree to disagree. More technically, there may be fixed points of the "group network" in which agents have different beliefs about specific features. Instead, we require each member of the group to reach, by repeatedly adjusting her own current interpretation to those of others, an acceptable individual interpretation that balances the pressures coming from her own internal mental states with those from others' interpretations– a kind of collective reflective equilibrium (Goodman 1955, Rawls 1971). Nonetheless, shared recognition, in which all agents reach *the same* interpretation, is an important type of collective recognition that plays a major role in our subsequent analysis.

A model of collective recognition should address a few fundamental questions. The first, most obvious question is the "existence problem:" can collective recognition as defined above be achieved? Without a positive answer, our modeling effort would be meaningless. The second question is about "creativity:" does a collective recognition have to reflect a pre-existing interpretation of at least one agent in the group, or can genuinely new interpretations or recognitions emerge from communication among agents? If new recognitions can emerge, a group-level analysis can go beyond merely investigating which individual interpretation might eventually prevail. The third question concerns the potentially dysfunctional aspects of group activity, which can generate conformity and judgmental arbitrariness – a tendency to unanimity that overrides the goal of realistically appraising situations (Janis, 1972). Can the model express such "groupthink" or related pathologies? Below, we provide a general, "closed form" answer to these questions.

We rephrase the "existence" question in terms of whether a network of interacting associative memories can preserve the "fixed point properties" of individual associative memories (i.e., whether the group can converge to a stable collective recognition of the environmental input). The answer is positive, and we express it as:

Proposition 1 (existence). *Given a network of associative memories (s, W) and the update rule (2), there is always at least one collective recognition for any input received by the agents.*

The formal proof is elaborate, and can be found in Appendix 2. An intuitive explanation can, however, be outlined by referencing Figure 6 and expression (2): the group model has exactly the structure of a Hopfield network, with the usual update rule extended to include other-agents' nodes weighted by communication connections. In other words, the group is a "collective associative memory" that merges the individual ones and adds between-agent interactions among nodes to within-agent ones. It follows that the group model inherits the main properties of the individual model, especially the existence and local stability of fixed points, which act literally as group memory states. Such memory states might not reflect agreement among group members, but their existence guarantees that by adjusting some individual interpretations to others' via communication, the group will achieve a collective recognition.

We now move to the "creativity" question. Can genuinely novel interpretations emerge out of agents' communication efforts? Consider two extreme cases. First, consider agents who are fully cognitively homogenous (i.e., their memories store the same set of situations). Second, consider agents who are fully heterogeneous (i.e., there are no overlapping situations stored in different agents' memories). In the first case, agents who perceive the same input from the environment do not need to communicate in order to converge on the same memorized situation because the recognition outcome (e.g., the fixed point toward which the recognition will converge) depends entirely on the dynamics of individual memories. In the second case, agents

will be able to reach an agreement only by some form of communication, which in this case is essential to collective recognition. Given our interest in how communication affects a group's recognition effort, we focus here on the latter case.[9]

In situations of full heterogeneity, the group can reach an agreement (a stable shared "state of mind") in one of two possible ways: it can associate the new reality with one experienced by a group member, or it can interpret the novel reality as a novel situation that does not fully match any of the group members' experiences. In this case, the group would generate an entirely new, shared state of mind. Below, we show how communication can induce a genuinely new shared state of mind, thereby demonstrating the potential "creative" effects of collective recognition efforts relative to individual ones.

The starting point is expression (2) above, which defines how each node in an individual's memory is updated given the states of that agent's own nodes and the state of others' nodes as mediated by communication. For a given shared recognition **s*** to be a fixed point, it must be true that, for all agents and all feature-nodes:

$$(3) \quad s *_i^k = \text{sgn}\left\{ \sum_j w_{ij}^k s *_j^k + \gamma \sum_{\substack{p=1 \\ p \neq k}}^{n} s *_i^p \right\}$$

When this condition is satisfied, no single node can deviate from its current state. Now, if without communication (or with $\gamma = 0$, which is the same) **s*** is not a fixed point, it must be true for at least one feature-node for each agent that:

$$(4) \quad \text{sgn}\left\{ \sum_j w_{ij}^k s *_j^k + \gamma \sum_{\substack{p=1 \\ p \neq k}}^{n} s *_i^p \right\} \neq \text{sgn}\left\{ \sum_j w_{ij}^k s *_j^k \right\}$$

(otherwise **s*** would be a fixed point even with $\gamma = 0$). Because **s*** is a shared recognition, $s*_i$ is the same for each agent; consequently (4) becomes:

---

[9] Most real-life situations lie between these two extremes. We analyze more realistic settings through a simulation in the next section.

$$\text{(4b)} \quad \text{sgn}\left\{\sum_j w_{ij}^k s*_j^k + \gamma(n-1)s*_i\right\} \neq \text{sgn}\left\{\sum_j w_{ij}^k s*_j^k\right\}$$

This in turn can be decomposed into two conditions:

$$\text{(5)} \quad \text{sgn}\left\{\gamma(n-1)s*_i\right\} \neq \text{sgn}\left\{\sum_j w_{ij}^k s*_j^k\right\}$$

and

$$\text{(6)} \quad \left|\gamma(n-1)s*_i\right| > \left|\sum_j w_{ij}^k s*_j^k\right|$$

(5) is implied by the fact that, without communication, **s*** is not a fixed point. Because $\sum_j w_{ij}^k s*_j^k$ is always

finite, it follows that one can always choose a γ large enough to verify (6). By applying this reasoning

repeatedly for each agent and each feature-node, it will be proved that:

Proposition 2 (creativity). *If γ is large enough, there will always be a shared recognition that does not*

correspond to any of the situations stored in individual memories.

Proposition 2 suggests how communication can induce a new stable state of mind in individual agents who

represent situations they would not conceive of if communication was absent. It indicates a genuinely creative

process for generating new interpretations of the environment. Communication can "force" an individual to

break the internal consistency of her mental states (induced by the connections within her own memory) to

establish a new interpretation that will account for the weight of others' hypotheses. When communication

weights are strong enough, they will lock the new mental state and make it stable. Thus, new stable states of

mind will arise from the recombination of different individual hypotheses. Critical to this process is a group's

degree of internal influence: groups whose members considerably influence each others' interpretations of

external inputs can end up recognizing new situations as truly different from anything they experienced before --

a whole new truth is created out of partial ones. Ironically, in this case, re-cognition may lead to new cognitions.

Yet, this property also implies that beyond a critical treshold of influence, communication can have deleterious effects. The formal argument goes as follows. For a given agreed interpretation $\mathbf{s}^{\alpha}$, let us call γ^{α} the minimal value of γ for which $\mathbf{s}^{\alpha}$ is a fixed point. This value will always exist since the "internal" weights of each individual memory are finite, and thus can always be overridden by the effects of communication. In a way, there is no unlimited individual stubbornness. This reasoning can be pushed further. It suffices to pick a γ^{ω} that is the max of the set of the γ^{α} for each possible shared recognition to prove that:

Proposition 3 (credulity). *Provided that γ is large enough, ANY arbitrary shared recognition can be a fixed point.*

High-influence groups tend to be increasingly "credulous," prone to agree on everything - a state reminiscent of the pathologies of groupthink. This result is significantly reinforced by a further proposition:

Proposition 4 (consensus). *Provided that γ is large enough, ALL fixed points must be shared recognitions.*

Proposition 4 establishes that, as communicative pressure increases, only conformist outcomes are possible: only perfect consensus can be stably sustained. The proof is only slightly more involved. Consider a fixed point that is *not* a shared recognition. Then, in light of (4), we must consider two different cases:

a) For some agent k and some feature i, the majority of the other agents give i a sign opposite the one assigned to it by agent k. Thus:

$$\text{sgn}(s*_i^k) \neq \text{sgn}\left(\gamma \sum_{\substack{p=1 \\ p \neq k}}^{n} s*_i^p \right)$$

If this is true, given (5), there will always be a γ for which this implies that

$$s *^k_i = -s *^k_i$$

which of course is incorrect;

b) For some agent k and some feature i, the number of other agents that give i a sign opposite the one assigned to it by agent k equals the number of other agents that give i the same sign as the one assigned to it by agent k. Thus:

$$\gamma \sum_{\substack{p=1 \\ p \neq k}}^{n} s *^p_i = 0$$

In this case, whatever the value that $s *^k_i$ takes, that value will decide what the majority interpretation of feature i is; consequently, there will always be (n-1)/2 agents other than k (q≠k) who will find themselves in the "minority" position analyzed sub a). As we have seen, for a large enough γ, this condition would lead to $s *^q_i = -s *^q_i$. Thus, provided that γ is large enough, there can be no fixed points which are not shared recognitions. Besides the details of the proof, the interpretation of Proposition 4 is clear: when a group is in a very high influence condition, no individual agent will put her opinion against that of the majority in her group. Consensus overrides realism.

In sum, propositions 2 through 4 suggest that an increase in groups' internal influence has both a positive and a negative effect on recognition outcomes. On the positive side, the creative properties of communication intensity (proposition 2) guarantee that genuinely new situations can be identified as such. If all agents' representations of the environment are partially wrong, intense communication may help generate a new recognition that is closer to reality by compensating for individual errors with the collective wisdom that results from the composition of the right hypotheses, which are diffused among different agents. If communication was unable to induce new fixed points, this generation would be impossible (we exploit this property to discuss the noise-filtering effects of communication in section 4.2). On the negative side, as groups' internal influence surpasses a critical threshold, the pressure to conform can push the group to converge to any arbitrary

configuration. Thus, the positive effect of communication on group performance can be progressively eroded by conformism.

Communication Structures, Division of Labor, and Recognition of Noisy Environments – A Simulation Analysis

Above, we established some basic properties of collective recognition. These properties are obtained in a very stylized, abstract setting. We now refine our model to capture some select aspects of more realistic settings. For instance, as we noted at the outset, analogical interpretation is likely to be important in novel environments, in which agents are structurally ignorant and information about the environment may be unreliable, distorted, incomplete, or even inconsistent. Further, agents may vary in their level of specialization, and how they communicate can be quite different across settings. The goal of this refinement is not to analyze comprehensively the wide set of contingencies that might be relevant to collective recognition. Such analysis is well beyond the scope of this paper. We view our refinement and associated analysis as a first, meaningful step in this direction, one that hopefully demonstrates the usefulness of this line of inquiry.

We refine the baseline model in three ways. First, we introduce noise. Agents perceive the environment more or less inaccurately, so that they always perceive states of the world that are different from the ones they already know, and different agents may hold different perceptions of the same, "true" state of the world. Second, agents have limited knowledge of the state space, and are initially unable to completely discriminate between alternative states of the world due to division of labor. Agents with different experiences and roles are inclined to pay attention to some features of the world, and are unable to make diagnostic use of others (Dearborn and Simon, 1958). Third, we consider alternative communication structures. Within this framework, we consider how different communication structures affect the accuracy and speed of collective recognition. For instance, is

a centralized structure of communication more or less effective than a decentralized one is in noisy environments?

Augmented realism is costly. First, it requires us to move from closed-form analyses to computer simulations. Second, it implies the introduction of additional parameters, thereby creating a large parameter space, which we can only sample from rather than explore exhaustively. Out of the many conceivable communication structures and cognitive divisions of labor, we consider only two elementary types, although our analysis of noise and communication intensity is more systematic.

Setup. Building on the structure we introduced above, our simulation model has three building blocks: an environment of situations; agents as associative memories; and communication structures among them. We describe them in turn.

Environment. The environment is represented as a set of situations (cfr. §2), each described by 30 binary features. We consider 5 situations (that represent true states of the world) and designate them as equidistant in terms of Hamming (bit by bit) distance.[10] The agents receive signals from the environment, which can be more or less *noisy*. Given the binary nature of the individual features, we model the effect of noise as a change in sign of the feature coding. So, if feature i has a "true" state -1, noise will make it be perceived as being in state 1. The level of noise depends on the probability that a given feature changes its state. For example, setting the noise parameter to prob = 0.2 translates into a 20% probability that any feature is perceived with the "wrong" sign.

Agents. Each agent is modeled as an associative memory who stores a limited number of situations (in our model, the agent stores 5 situations), each consisting of 30 features. Each agent stores different situations. We represent cognitive heterogeneity and the associated division of cognitive labor by assuming that each agent can discriminate states of the world in regard to a limited number of features, and the set of such features varies by agent according to her specialization. We express such "specialized blindness" by having each agent's stored situations differ only by a limited number of features, the ones reflecting her specialized focus. The remaining

---

[10] It can be shown that there is no loss of generality in assuming equidistance.

features have the same value in all memorized situations, and thus do not help discriminate states of the world. In our model, agents are "blinded" over 3/5 of the features. We also tested a 2/5 ratio, with similar results. In addition to "specialists," we introduce "generalists" for whom all features can be discriminating.

Group Structure. We consider the full and star structures introduced above. For each, we consider 4 agents, three specialists and a generalist. In the full communication structure, all the agents communicate with each other. In the star structure, there is a central agent, who is a generalist, and three peripheral ones, who are specialists, and who can communicate only with the central agent. Groups' degree of internal influence is modeled through γ as described above.

Simulation Plan. This basic setup allows us to study collective recognition efforts generated by alternative communication structures under different conditions of noise and communication intensity. Specifically, we consider a 2x8x30 simulation design, which has the 2 prototypical communication structures we just described, 8 levels of noise, and 26 levels of the γ parameter (in the 0-50 range). The noise parameter ranges between 0 (no noise) to 0.66 (66% probability that a node receives a random signal). The situations stored in agents' memory are the same in all of the 2x8x30 simulation conditions. For each cell of the simulation design, we generated 500 independent runs. For each run, a true state of the world is generated at time 0. Each agent receives an input corresponding to the true state of the environment, corrupted by noise. Noise is independent for each agent (i.e., there is no systematic relationship between how noise distorts agent i's and agent j's perception). Once the agents have initialized their state according to the input received, both memory and communication dynamics are activated. At each step of each run, a single node of a single agent is randomly selected for updating according to (4), reflecting both the constraints of each agents' internal memory and the state of corresponding nodes in other agents as mediated by communication links. The simulation stops after there are no changes in the state of the network for 150 consecutive steps. At that point, the interpretation is considered stable.

Results. We consider two focal performance metrics: the recognition error and the time to recognition. We express the former in terms of the sum of the group's erroneous bits once a stable recognition is reached, and the latter as the number of steps that are needed to reach such an outcome. We first comment on properties emerging from the simulation that are common to both structures. We then focus on how differences in structure affect such basic properties.

a. Convergence to shared interpretations as a result of communication intensity. The effects of communication intensity on collective recognition are most visible when noise is absent. As Figure 7 a-c shows, there is a high, stable level of error for low levels of communication intensity because agents tend to relax to their own memorized situation, thereby producing an erroneous interpretation: due to the division of cognitive labor, agents' interpretations of the environment are only partially correct. As the intensity of communication increases, however, there is a rapid shift to total agreement on the correct interpretation. Communication helps assemble partial correct interpretations, thereby producing stable new states of mind in the individual agent. Not surprisingly, in transparent worlds, communication has decisive effects.

b. Collective filtering of noise. As noise is introduced, the group must also filter out the possible distortions induced by it. Figure 7 a-c shows the following properties. First, communication filters noise: it reduces collective errors, although imperfectly (even with low levels of noise, some residual error persists). The basic intuition is that if there is at least some independence in the noise perceived by agents, the fact that multiple agents receive information about the same feature builds redundancy. Redundancy in information channels is exploited by communication: more observers have more chances in the aggregate to observe the right signal than a single observer does, and communication allows individual misperceptions to be corrected. Second, as noise increases, the ability of communication to correct distorted inputs diminishes. Importantly, this effect (particularly for high levels of communication) occurs not because agents develop diverging interpretations.

Rather, they increasingly converge to a wrong interpretation, thereby fitting noise. Table 1 shows how the ratio

for cases of any type of convergence relative to cases of correct convergence changes only as γ increases.

noise	$\gamma=10$	$\gamma=40$
0.13	0.81	0.69
0.27	0,35	0,16
0.40	0,18	0,1
0.53	0	0,003

Table 1: Ratio of correct to incorrect convergences

This result can be better understood in light of properties P2, which allows convergence to new patterns, and

P3, which says that any pattern can be a fixed point, as demonstrated above. In particular, it reflects the tendency

to converge to arbitrary patterns, which increases with γ (credulity). Whether coordination on the wrong beliefs

is better than no coordination at all is highly task-dependent – we do not deal here with this issue.

Third, the effects of communication on noise reduction are non-monotonic. That is, for high levels of noise,

low levels of communication intensity produce more error than no communication does. Once a critical level of

communication intensity is achieved, error decreases. This counterintuitive result can be understood by

reference to equation (4), which expresses the joint effect of memory and communication on individual nodes.

When γ is not high enough to systematically correct for noise, it can still be high enough to accidentally alter the

value of the states of single features. It thereby acts in ways similar to noise. Thus, under the effect of

communication, agents may wander in the feature space without converging to the right belief; occasionally this

wandering may lead agents out of the original path and toward some neighboring basin of attraction.

All of these properties can be extended to the time needed to reach a stable state (time-to-recognition).

Communication intensity reduces the time needed to reach a stable state; it does so less effectively as noise

increases; and it does so non-monotonically (see Figure 8 b-d). When γ increases, agents are more prone to

align their beliefs with those that are held by their group members. Noise increases distances between individual

perceptions and thus slows down convergence. Finally, the "wandering" effects of low γ affect time-to-

recognition, as well as the error rate, for the same reasons. In fact, the effect is even stronger because in all cases in which wandering does not deviate, trajectories from the original basin of attraction are not reflected in error rates, but instead affect convergence speed.

c. Comparing structures:

Although some comparative patterns clearly emerge in Figure 7 a-d, they can be better captured by simple performance ratios for the full combination of parameters. Figure 8 a-b represents the ratio between the collective error and the time-to-recognition of the fully-connected and star structures, respectively. In both cases, ratio values below 1 indicate that the fully-connected structure performs better, and values above 1 indicate the opposite.

Intuitively, the advantages of increasing communication intensity are stronger in structures that, keeping the number of agents constant, have denser communication networks. Fig. 8a shows that in a vast region characterized by high values of noise and communication intensity, the fully-connected structure enjoys a considerable error reduction advantage over the star structure. This advantage has at least two causes. First, the fully-connected structure is denser (there are more connections), which produces a pure multiplicative effect over γ. Second, each node in the fully-connected structure receives messages from a larger number of information sources (redundancy). This tendency increases the probability that a node will be reached by the "right" message. In contrast, in the corner characterized by low values of both parameters, the star structure enjoys a clear advantage. This result reflects the non-monotonicity of communication intensity effects, which are more marked in the communication-dense fully-connected structure. Further, another mild form of non-monotonicity can be observed in the opposite corner of the parameter space. The relative superiority of the fully-connected structure decreases for high noise and high communication intensity. This behavior reflects the fact that the denser communication network of the fully-connected structure exhausts the beneficial effects of

increasing γ more rapidly than does the star structure, which slowly catches up in performance as a consequence.

We now consider time-to-recognition. Although the fully-connected structure maintains its basic superiority over a large area of the parameter space, there are important differences compared to the error-reduction performance. First, the corner in which the star structure enjoyed an error-reduction advantage also shows a time-to-recognition advantage. The corner is now broader, however, and the advantages more pronounced. Second, in the area in which the fully-connected structure enjoys superior error-reduction performance, the advantage in time-to-interpretation is considerably less marked, and it depends less on noise. Third, the star structure's performance catches up (for high levels of γ) more rapidly to that of the fully-connected structure for all levels of noise. Until now, we have not considered how the cost of communication affects the relative performance of the two communication structures. Yet a major motivation for star structures is that they significantly economize on communication costs. As is well known, a fully connected network of n agents has $n \times (n-1)/2$ communication links, but a star has only *n-1* links. Because our model considers only the interpretive stage of decision-making, a direct cost/benefit comparison cannot be easily performed. Nonetheless, a simple qualitative argument suggests that as the cost of communication increases, not only will the corner of the parameter space in which the star is superior be enlarged, but the opposite corner might also "emerge," giving rise to an interesting re-switch of structures due to the non-monotonicity phenomena noted above: the star might turn out to be superior not only for low noise and low communication intensity, but also for very high communication intensity. Furthermore, this process is bound to be more relevant to the extent that the speed of interpretation is more valuable than accuracy is.

DISCUSSION AND CONCLUSION

In novel worlds, strategic choice is often premised on analogy, and recognition lies at its heart. Yet despite recognition's centrality, we have only limited knowledge of its nature and its influence on strategic choice in

individuals, much less in the multi-agent decision-making settings in which these choices typically occur. In this paper, we begin to chart this territory, and offer three contributions.

First, we derive a formal platform that plausibly captures recognition processes in groups of decision-makers. Our map of this uncharted territory narrows attention to three pivotal coordinates. Most obviously, we propose that strategic choice in novel and structurally obscure environments can be fruitfully represented as recognition of the new in terms of the old. In addition, we put forth a conception of individuals as memories, and of groups as aggregates of interacting memories. Further, we conceive the linkages across such memories in terms of how readily individuals within a group influence each other through communication. All in all, our map offers a novel perspective on decision-making in novel worlds and group processes. Karl Weick (1990) reminds us that when no map is available, any map is useful, regardless of how accurate it is. Although behavioral realism drives our effort to develop a map that is as accurate as possible, we are sympathetic with Weick's remarks. With just a few exceptions, which have focused mainly on individual rather than group processes (e.g., Gavetti, Levinthal, and Rivkin, 2005), analogical strategic choice is uncharted territory. Whatever its imperfections, we believe our map offers a useful guide to a better understanding of the reality of strategic choice.

Second, we use this formal structure to derive some basic properties of collective recognition in novel worlds. Our analyses highlight how group-level characteristics affect recognition outcomes. The literature on groups and teams has focused on many parameters of group processes and outcomes (e.g., Gruenfeld et al, 1996; Sundstrom, Busby, and Bobrow, 1997; Phillips, Northcraft, and Neale, 2006), but has not examined how analogy and recognition processes affect group decision-making. Our study posits a mechanism for transmitting influence across a network, and thus suggests how diverse individuals whose understanding of a problem is each limited and often mutually incongruent can "share" their beliefs and thereby converge toward a collective interpretation (Gruenfeld et al, 1996; Woolley et al, 2007). It also shows analytically how variables of

communication, such as its influence, intensity, and structure, affect group outcomes. Such findings are especially important to settings where groups are highly collaborative (Edmondson, Bohmer, and Pisano, 2001; Reagans, Argote, and Brooks, 2005). More fundamentally, our work is consistent with recent work in this domain that takes individual-level neural processes as the central primitive to shed new light on groups' functioning (Woolley et al., 2007).

In particular, we show, in closed form, that communication can induce agents to converge to shared recognitions that are new to each of them. Groups of decision-makers that operate in analogical terms and establish a culture of strong internal influence are more likely to converge on shared interpretations that do not correspond to any of their members' prior experiences. This property suggests a crucial difference between the recognition efforts of individuals and teams of decision-makers. At the individual level, associative memory associates the novel reality with some memorized situation. In contrast, teams of decision-makers can be more creative: collective associative processes can lead to recognitions of novel realities that transcend the boundaries of individual memories. This property evokes an image of the group as operating in autonomy from its individual constituents, as if it was itself an individual who behaves according to her own (e.g. group-level) rules, generates her own outcomes, and even feeds back into the cognition of her individual constituents. Beyond these aspects, this property plays a particularly salient role when novelty constitutes a challenge for which past experience is insufficient. The notion that teams can be more creative than individuals are is not new (e.g., Hurst, Rush, and White, 1989), but our model suggests that the strength of this effect, particularly in groups of analogizers, is deeply affected by the group's level of "*internal influence.*" This property finds some analogues in DeMarzo et al's (2003) model of opinion formation in social settings. Similar to our model, DeMarzo et al's model suggests that, in social contexts such as groups, "the beliefs of all agents converge over time to a weighted average of initial beliefs." (DeMarzo et al., 2003: 913) Although DeMarzo et al develop a model that does not explicitly address recognition and its role in novel settings, their results are strikingly

consistent with ours. Despite these virtues, "*internal influence*" can take its toll. As the weight of communication increases, the group becomes more vulnerable to groupthink, which in turn may make interpretations largely arbitrary.

We then tailor the model to explore, in simulation form, more specific conditions that better approximate the complications of real decision-making. Because the parameter space corresponding to this more realistic setting is expansive, we analyze only a small portion of it. In particular, we consider groups of agents that are affected by the cognitive blinders of division of labor and ask whether they can do better than individuals do in noisy worlds, where the information they receive is distorted and unrealiable. We show that communication can correct for the errors induced by the division of labor, and, pehaps more importantly, can filter the effects of noise. The effect of communication on the quality of recognition is, however, non-monotonic. Given groups' tendency to fall into a conformist mode when internal influence is high, noise-fitting prevails over noise-filtering in these situations. We also compare the effects of alternative group structures on recognition outcomes, and find that groups' communication density plays a central role. Different group structures imply different levels of communication density, and, *ceteris paribus,* higher levels of density amplify our results of the benefits and dysfunctions of communication on recognition outcomes.

Beyond the specific details of our results, the simulation demonstrates that recognition outcomes are extremely sensitive to the group's cognitive composition, communication structure, and intensity. We take this overarching pattern to delineate a fruitful path for future research, which we believe needs to couple a more extensive exploration of our model's parameter space with experimental work. This research should have two main objectives. First, it should assess whether the model's core predictions are empirically supported. In addition, it would be useful to replicate the gist of the simulation exercise via human-subject experiments. This line of inquiry would allow the model's parameters to be calibrated more precisely vis-à-vis the outcomes of

interest, which is essential given the extreme sensitivity of recognition outcomes to such parameters. Indeed, we are currently engaged in a follow-up study that proceeds along these lines.

Third, more prospectively, we believe this work can open a genuinely generative dialogue among theories of choice and action that, although not generally associated with strategic choice, emphasize similarity matching as the core cognitive mechanism underlying choice. We refer in particular to March and colleagues' logic of appropriateness (cfr. March, Schulz, and Zhou, 2000, and March and Olsen, 2006), and Gilboa and Schmeidler's case-based decision-theory (cfr. Gilboa and Schmeidler, 2001). Our formal effort and these theories have a lot to offer each other. Thus, opening a dialogue among them would benefit our understanding of choice, particularly strategic choice, in novel worlds.

The logic of appropriateness applies when individuals and groups face uncertain situations. When they do, "the processes of reasoning are not primarily connected to the anticipation of future consequences (...) Actors use criteria of similarity and congruence, rather than likelihood and value." (March and Olsen, 2006: 690) According to this view, agents choose by matching choices and behaviors to a given situation, and the repertoires of choices are idiosyncratic to individuals' roles and identities in that situation. Decision-making is thus rule-based (March, Schulz, and Zhou, 2000); decision and action, in this view, are driven by programs, standard operating procedures, and routines (Nelson and Winter, 1982; March and Simon, 1993; March and Olsen, 2006). This perspective adds to our approach the idea that repertoires of behavior (in our language, the experiences that individuals store in their memories) and their rules of appropriateness are tied to individuals' roles and identites, which in turn reflect institutional logics and memories. Our effort, on the other hand, is a first attempt to formally analyze the cognitive underpinnings of recognition at both the individual and collective level, thereby illuminating how repertoires are matched to novel situations. Although it captures only the essentials of a remarkably complex process, it is rich enough to show that the analogy-driven collective recognition of current situations is not based on a crude direct matching of present problems to stored

repertoires, but is instead mediated by processes of memory recall and communication that can significantly transform the original repertoires and generate new interpretations. We thus believe our perspective can both enrich rule-based notions of choice and contribute to a growing stream of theory about the mechanisms of routine formation and performance (Cohen and Bacdayan, 1994; Edmondson, Bohmer, and Pisano, 2001; Feldman and Pentland, 2003; Pentland and Feldman, 2005).

Case-based decision theory (CBDT) provides a more utilitarian, and formal, counterpart to the logic of appropriateness, but it rests on a similar logic. It is conceived to model how agents decide when they face problems for which they do not know the state space and its associated outcomes (structural ignorance). The guiding intuition is that, in such contexts, "the main reasoning technique that people use is drawing analogies between past cases and the one at hand." (Gilboa and Schmeidler, 1995: 608). Gilboa and Schmeidler provide powerful formal foundations of a theory of individual decision-making with limited memory and reasoning by analogy. We view our model as complementary to CBDT for two reasons. On the one hand, Gilboa and Schmeidler offer a parsimonious full-fledged model of analogical decision-making, while we focus exclusively on the recognition aspect of such processes. On the other hand, we suggest a more plausible and efficient mechanism for implementing search in memory. Instead of the systematic scanning of memory records that CBDT implies, we model content-addressable memories that look at the content of present input stimuli to converge directly to the most similar memory template. Further, we suggest ways to move from individual to collective analogy-based decision making.

Finally, in addition to these connections with existing theories of choice and action, we wish to point out parallels between our account of how recognition occurs in groups and that of recent work on the emergence of organizational forms. For instance, Hannan, Pólos, and Carroll (2007) consider audiences that are external to focal organizations as playing a decisive role in whether emergence occurs. They contend that "many categories and forms might have originated in fairly structured situations, those in which the audience has a well-developed

language (including schemata for categories). In these kinds of situations, new categories can be formed by analogy to existing ones (by making slight modifications in schemata), by merging a pair of categories, and so forth" (Hannan, Pólos, and Carroll, 2007: 37-38). In their account, as in ours, interaction among agents that are involved in what are essentially recognition processes is crucial to the phenomenon in question. Further, we believe recognition transpires even in the less-structured situations that this literature considers. When actors develop a new code to assess whether a set of organizations might cohere into a category, as beer enthusiasts did when they changed the code for breweries to reflect the type of producer as well as the product itself (Hannan, Pólos, and Carroll, 2007; cf. Carroll and Swaminathan, 2000), they typically do not develop these codes *ex novo.* More often than not, they use past memories of possible codes to recognize which ones are relevant to the situation at hand. Our model was not originally designed to represent the emergence of organizational forms. Nonetheless, the cognitive phenomenology captured by our model is sufficiently aligned with what is thought to underlie forms' emergence that we believe it can be tailored to contribute meaningfully to this debate.

To conclude, we provide theoretical foundations for a largely understudied form of strategic choice. We believe our theoretical foundations, especially if coupled with theoretical approaches of a similar bent, promise to chart a territory that is fundamental to our understanding and improvement of strategic choice.

Bibliography

Amit, D.J., N. Brumel, M.V. Tsodyks. 1994. Correlations of cortical Hebbian reverberations: Theory versus experiment. Journal of Neuroscience, 14: 6433-6445.

Anderson, J.R. 1995. Learning and Memory: An Integrated Approach. New York: Wiley

Anderson, J.R., G.H. Bower. 1980. Human Associative Memory. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brandenburger, A., H. Stuart. 1996. Value-based business strategy. Journal of Economics and Management Strategy, 5: 5-24.

Cantambrone, R. 2002. The effects of surface and structural feature matches on the access of story analogs. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28: 318-334.

Carroll, G.R., A. Swaminathan. 2000. Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry. American Journal of Sociology, 106: 715-762.

Cohen, M.D, P. Bacdayan. 1994. Organizational routines are stored as procedural memory: evidence from a laboratory study. Organization Science 5: 554-568.

Cyert, R.M., J.G. March. 1963. A Behavioral Theory of the Firm. Englewood Cliffs, NJ: Prentice-Hall.

Camerer, C., D. Lovallo. 1999. Overconfidence and excess entry: An experimental approach. American Economic Review, 89: 306-318.

Dearborn, D.C., H.A. Simon. 1958. Selective perception: A note on the departmental identification of executives. Sociometry, 21: 140-144.

DeMarzo, P.M., D. Vayanos, J. Zweibel. 2003. Persuasion bias, social influence, and unidimensional opinions. Quarterly Journal of Economics 118: 909-968.

Denrell, J., C. Fang, S.G. Winter. 2003. The economics of strategic opportunity. Strategic Management Journal, 24: 977-990.

Edmondson, A.C., R.M. Bohmer, G.P. Pisano. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. Administrative Science Quarterly, 46: 685-716.

Feldman, M.S., B.T. Pentland. 2003. Reconceptualizing Organizational Routines as a Source of Flexibility and Change. Administrative Science Quarterly, 48: 94-118.

Fuster, J.M. 1995. Memory in the Cerebral Cortex. Cambridge, MA: MIT Press.

Gavetti, G., and D. Levinthal. 2000. Looking forward and looking backward: cognitive and experiential search. Administrative Science Quarterly, 45: 113-137.

Gavetti, G., D. Levinthal, and J.W. Rivkin. 2005. Strategy-making in novel and complex worlds: the power of analogy. Strategic Management Journal, 26: 691-712

Gavetti, G., J.W. Rivkin. 2007. On the origin of strategy: Action and cognition over time. Organization Science, 18: 420-439.

Gavetti, G., J.W. Rivkin, E. Johnson. 2002. Lycos (A): The Tripod Decision. Harvard Business School Case 9-702-435. Boston, MA.

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7: 155-170.

Ghemawat, P 1999. Strategy and the Business Landscape. Reading, MA: Addison-Wesley.

Gick, M.L., K.J. Holyoak. 1980. Analogical problem solving. Cognitive Psychology, 12: 306-355.

Gilboa, I., D. Schmeidler. 1995. Case-based decision theory. Quarterly Journal of Economics, 110: 605-639.

Gilboa, I., D. Schmeidler. 2000. Case-based knowledge and Induction. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 30: 85-95.

Gilboa, I., D. Schmeidler. 2001. A Theory of Case-Based Decisions. Cambridge: Cambridge University Press.

Goodman, N. 1955. Fact, Fiction, and Forecast. Harvard University Press, Cambridge, MA.

Gruenfeld, D.H., E.A. Mannix, K.Y. Williams, M.A. Neale. 1996. Group composition and decision-making: How member familiarity and information distribution affect process and performance. Organizational Behavior and Human Decision Processes, 67: 1-15.

Hannan, M.T. L. Pólos, G.R. Carroll. 2007. Logics of Organization Theory. Princeton, NJ: Princeton University Press.

Hertz, J., A. Krogh, R. Palmer. 1991. Introduction to the theory of neural computation. Reading, MA: Addison Wesley.

Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79: 2554-2558.

Huff, A.S., J.O. Huff. 2000. When Firms Change Direction. New York: Oxford University Press.

Hummel, J.E., K.J. Holyoak. 2003. A symbolic-connectionist theory of relational inference and generalization. Psychological Review, 110: 220-265.

Hurst, D.K., J.C. Rush, R.E. White. 1989. Top management teams and organizational renewal. Strategic Management Journal, 10: 87-105.

Hutchins, E. 1995. Cognition in the Wild. Cambridge, MA: MIT Press.

Janis, I. 1972. Victims of Groupthink: A Psychological Study of Foreign Policy Decisions and Fiascoes. Boston, MA: Houghton-Mifflin.

Kokinov, B.N., A.A. Petrov. 2001. Integration of memory and reasoning in analogy-making: The AMBR model. In D. Gentner, K. Holyoak, and B. Kokinov, eds. The Analogical Mind: Perspectives from Cognitive Science: 59-124. Cambridge, MA: MIT Press.

Lant, T., Z. Shapira. 2000. Organizational Cognition: Computation and Interpretation. Hillsdale, NJ: Lawrence Erlbaum Associates.

Levinthal, D.A. 1997. Adaptation on Rugged Landscapes. Management Science, 43: 934-950.

March, J.G. 2006. Rationality, foolishness, and adaptive intelligence. Strategic Management Journal, 27: 201-214.

March, J.G., J.P. Olsen. 1976. Ambiguity and Choice in Organizations. Boston, MA: Scandinavian University Press.

March, J.G., J.P. Olsen. 2006. The logic of appropriateness. In M. Moran, M. Rein, R.E. Goodin (eds.), The Oxford Handbook of Political Science: 689-708. New York: Oxford University Press.

March, J.G., M. Schulz, X. Zhou. 2000. The Dynamics of Rules: Change in Organizational Codes. Stanford, CA: Stanford University Press.

March, J.G., H.A. Simon. 1993. Organizations, 2nd Ed. Cambridge, MA: Blackwell Business.

Marchiori, D., M. Warglien. 2005. Constructing shared interpretations in a team of intelligent agents: The effects of communication intensity and structure. In T. Terano, H. Kita, T. Kaneda, K. Arai, H. Deguchi (eds), Agent-Based Simulations: From Modeling Methodologies to Real-World Applications: 58-71. Berlin: Springer Verlag.

McRae, K., V.R. de Sa, M.S. Seidenburg. 1997. On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology, 126: 99-130.

Meindl, J., C. Stubbart, J.F. Porac. 1996. Cognition within and between Organizations. Sage: Thousand Oaks, CA.

Mintzberg, H. 1978. Patterns in strategy formation. Management Science, 24: 934-948.

Miyashita, Y. 1988. Neuronal correlate of visual associative memory in primate temporal cortex. Nature, 335: 817-820.

Nelson, R.R., S.G. Winter. 1982. An Evolutionary Theory of Economic Change. Cambridge, MA: Harvard University Press.

Neustadt, R.E., E.R. May. 1986. Thinking in Time: The Uses of History for Decision Makers. New York: Free Press.

Ocasio, W. 1997. Towards an attention-based view of the firm. Strategic Management Journal, 18: 187-206.

Pentland, B.T., M.S. Feldman. 2005. Organizational routines as a unit of analysis. Industrial and Corporate Change, 14: 793-815.

Phillips, K.W., G.B. Northcraft, M.A. Neale. 2006. Surface-level diversity and decision-making in groups: When does deep-level similarity help? Group Processes and Intergroup Relations, 9: 467-482.

Pinker, S., A. Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. Cognition, 28: 173-193.

Porter, M.E. 1996. What is strategy? Harvard Business Review, 74(6): 61-78.

Rawls, J. 1971. A Theory of Justice. Cambridge, MA: Harvard University Press.

Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. Management Science, 51: 689-701.

Rosch, E. 1978. Prototype classification and logical classification: The two systems. In E. Scholnik (ed.), New Trends in Cognitive Representation: Challenges to Piaget's Theory: 73-86. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., C.B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7: 573-605.

Rumelhart, D.E., A. Ortony. 1977. In R.C. Anderson, R.J. Spiro, W.E. Montague (eds.), The representation of knowledge in memory. Schooling and the Acquisition of Knowledge: 99-135. Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H.A. 1991. Bounded rationality and organizational choice. Organization Science, 2: 125-134.

Smolensky, P., G. Legendre. 2006. The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar. Cambridge, MA: MIT Press

Sundstrom, E., P.L. Busby, W.S. Bobrow. 1997. Group process and performance interpersonal behaviors and decision quality in group problem solving. Group Dynamics – Theory, Research, Practice, 1: 241-253/

Sutton, J. 2003. Memory. The Stanford Encyclopedia of Philosophy (Spring 2003 Edition), E.N. Zalta (ed.), at http://plato.stanford.edu/archives/spr2003/entries/memory/

Thagard, P. 1996. Mind: Introduction to Cognitive Science. Cambridge, MA: MIT Press

Thompson, L., D. Gentner, J. Loewenstein. 2000. Avoiding missed opportunities in managerial life: Analogical training more powerful than individual case training. Organizational Behavior and Human Decision Processes, 82: 60-75

Tversky, A. 1977. Features of similarity. Psychological Review, 84: 327-352.

Weick, K.E. 1990. Cartographic myths in organizations. In A.S. Huff (ed.), Mapping Strategic Thought: 1-10. London: Wiley.

Weick, K.E. 1995. Sensemaking in Organizations. Thousand Oaks, CA: Sage

Winter, S.G. 2000. The satisficing principle in capability learning. Strat. Management Journal. 21: 981-996.

Woolley, A.W., J.R. Hackman, T.E. Jerde, C.F. Chabris, S.L. Bennett., S.M. Kosslyn. 2007. Using brain-based measures to compose teams: How individual capabilities and team collaboration strategies jointly shape performance. Social Neuroscience, 2: 96-105.

FIGURE 1: ASSOCIATIVE RECALL: A VISUAL EXAMPLE

a   b
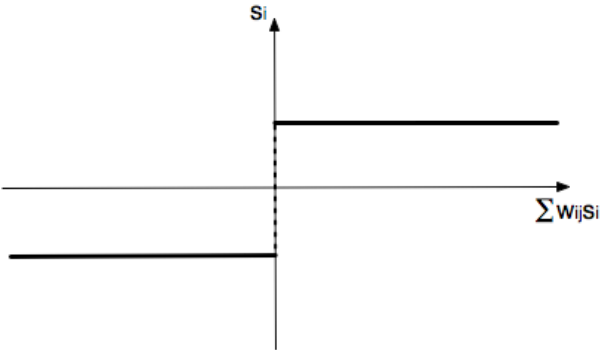
c   d   e
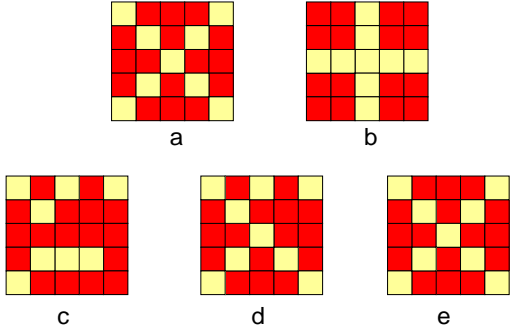
FIGURE 2: A NEURAL NETWORK

X_i

w_{ij}

X_j
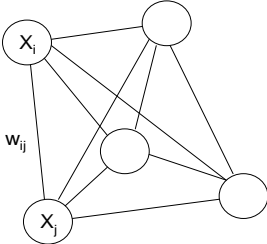
FIGURE 3: UPDATE RULE

S_i

$\sum W_{ij} S_i$

FIGURE 4: BASINS OF ATTRACTION

X = initial situation   ■**= memorized situation**
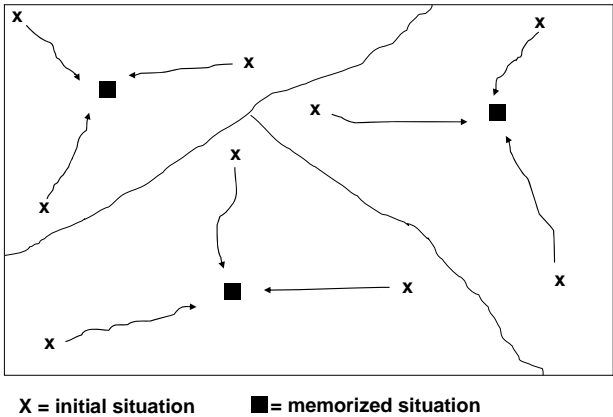
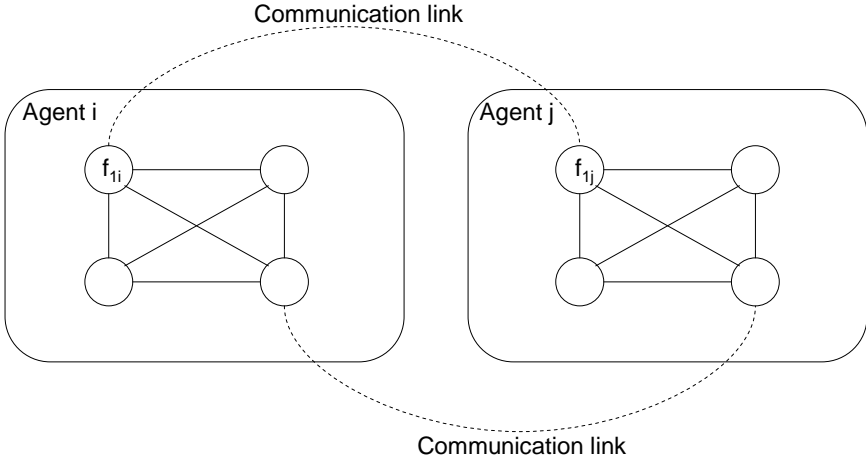FIGURE 5: INTRODUCING MULTIPLE AGENTS

Communication link

Agent i

f_{1i}

Agent j
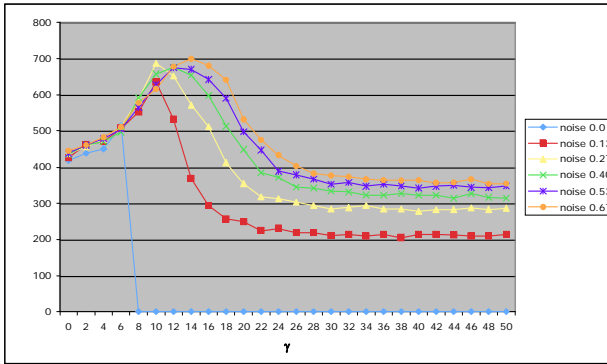
f_{1j}

Communication link

FIGURE 6: A MULTI-AGENT NEURAL NETWORK

S S^1
       S^2
       S^3

W^1	comm.	comm.
comm.	W^2	comm.
comm.	comm.	W^3

M
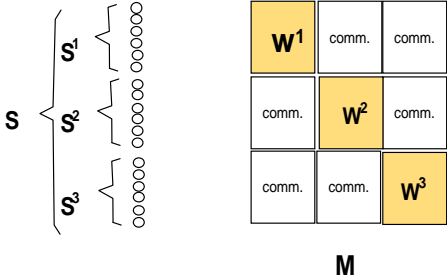
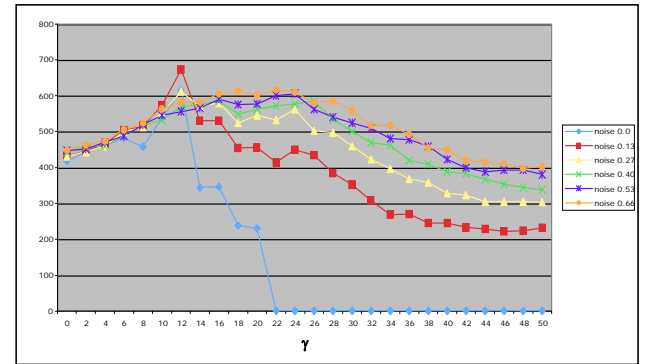FIGURE 7: SIMULATION RESULTS -- RECOGNITION ERROR AND TIME TO RECOGNITION
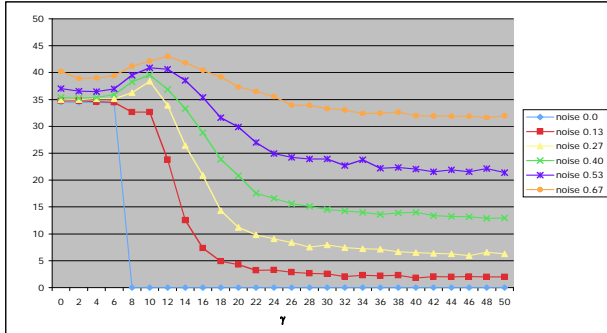
a: Team Error – Full Network
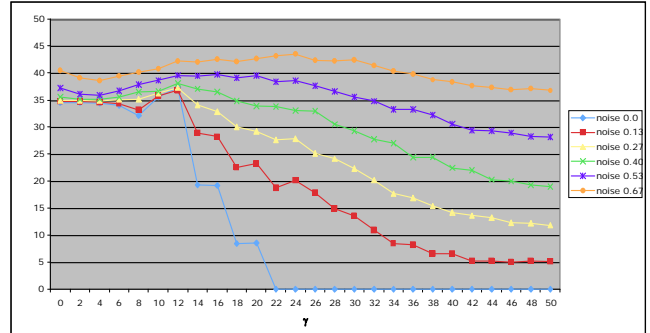
b: Team Error -- Star

c: Time to Recognition – Full Network

d: Time to Recognition – Star

TABLE 2: AVERAGE RATIO OF STANDARD DEVIATION TO AVERAGE OF SIMULATION RESULTS (OVER ALL γ VALUES)

Noise	Full network	Star
0.0		
0.13	0.98	0.53
0.27	0.56	0.37
0.4	0.38	0.26
0.53	0.26	0.20
0.67	0.19	0.16

a: Team Error

Noise	Full network	Star
0.0		
0.13	0.41	0.41
0.27	0.34	0.37
0.4	0.32	0.35
0.53	0.30	0.35
0.67	0.29	0.32

b: Time to Recognition

FIGURE 8A: SIMULATION RESULTS -- RECOGNITION ERROR (FULL/STAR)
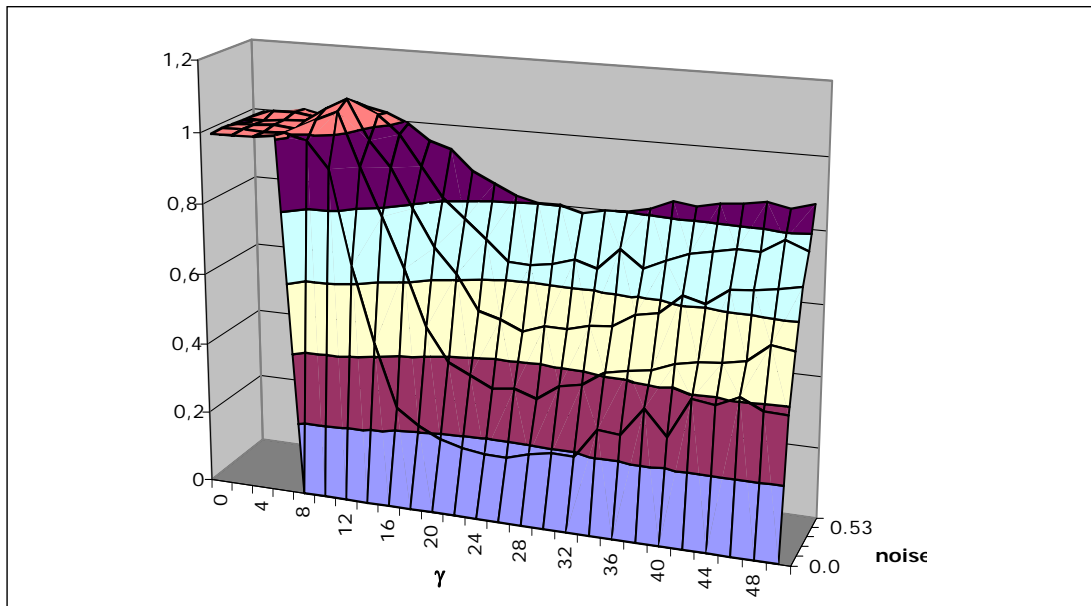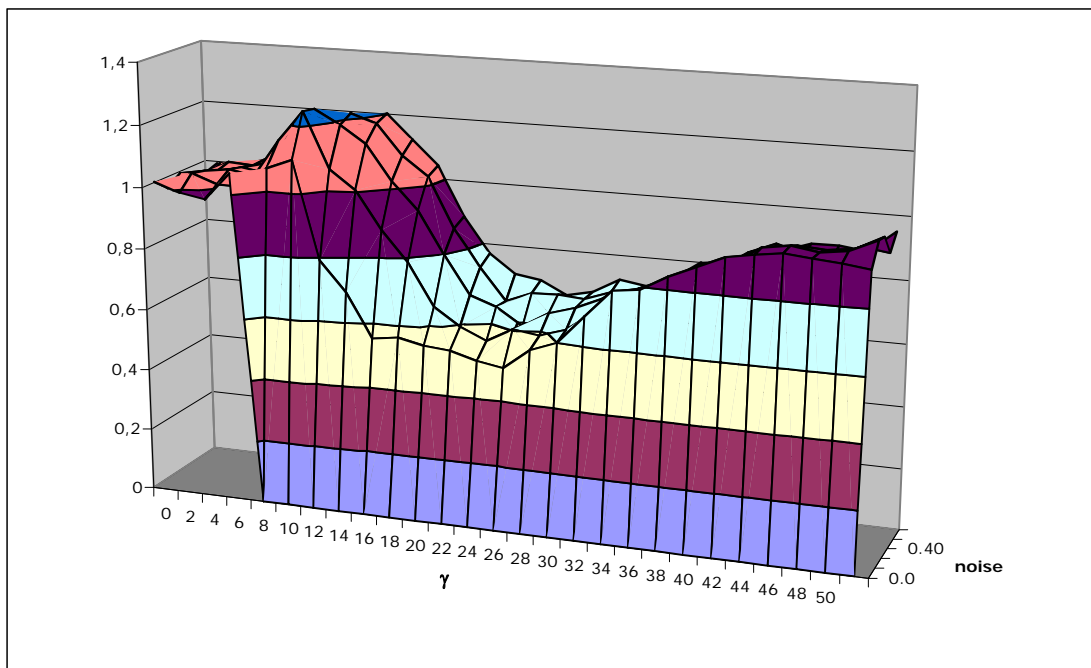

FIGURE 8B: SIMULATION RESULTS -- TIME TO CONVERGENCE (FULL/STAR)

APPENDIX 1. AN EXAMPLE OF THE MODEL'S MECHANICS

A simplified illustration of the network dynamics may help to show how the process of associative recall works.

A network of 4 nodes (s1,s2,s3,s4) has memorized a pattern (1, 1, 1,-1) through the following matrix W of connection weights:

$$W = \begin{vmatrix} .5 & .5 & .5 & -.5 \\ .5 & .5 & .5 & -.5 \\ .5 & .5 & .5 & -.5 \\ -.5 & -.5 & -.5 & -.5 \end{vmatrix}$$

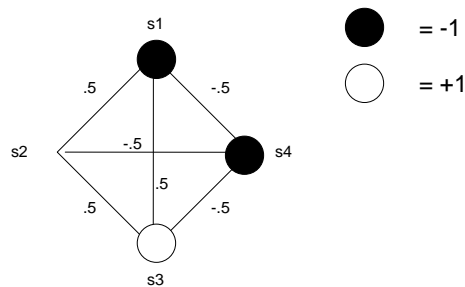Fig.A1 represents the network in an initial configuration (-1,1,1,-1)

Fig. A1

Imagine that node *s1* is randomly picked up (fig. A2). This node will "receive" from each other node sj an input given by the state of the "sending" node weighted by the connection weight w_{sj} between the former and the latter node. In the case of fig. A2:

$$(1 \times .5) + (1 \times .5) + (-1 \times -.5) = 1.5$$

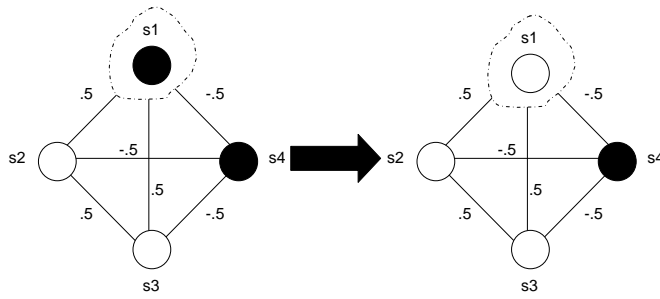The sign of the sum of the inputs is positive, so node *s1* will switch to a +1 state.

Fig. A2

Now node *s4* is randomly picked up (fig. A3). Once more, it will "receive" from the other nodes an input given by the state of the node as it is weighted by the connection weight (in the example of fig. A3):

$$(1\times\text{-}.5)+(1\times\text{-}.5)+(1\times\text{-}.5)=\text{-}1.5$$

The sign of the sum of the inputs is negative, so node *s4* will stay in a -1 state.
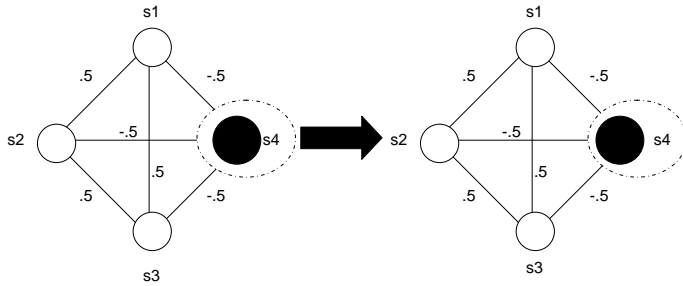

Fig. A3

It is easily verified that all nodes are now in a stable state, and that the memorized pattern (1, 1, 1,-1) has thus been associatively "retrieved" by the network

APPENDIX 2. PROOF OF PROPOSITION 1

The existence of fixed points for a nework of associative memories **(s, W)** and the update rule (2) can be demonstrated by extending the usual proof of the fixed-point properties of a Hopfield network (we follow closely the proof in Hertz, Krogh, and Palmer. 1991).

For notational simplicity, we will ignore the differences between agents by considering a generic vector of binary nodes **s**= (s₁, s₂, **...,** sₙ **),** which results from placing the different agents' nodes into an ordered sequence. In a similar vein, we will ignore the notational differences between connections within individual memories and communication links by considering a generic connection w_{ij} between the elements of **s,** which includes both kinds of connections.

We shall assume that nodes are not connected to themselves (i.e. w_{ii}=0), but the results also hold if the elements on the main diagonal of W are non-negative (indeed, it would make little sense to imagine hypotheses that are self-inhibiting!).

The fixed-point property will be demonstrated by showing that rule (2) is a monotonic, non-increasing function. Given that the set of possible configurations of **s** is finite, this condition trivially implies that at least one fixed point exists and there is convergence to it in finite time.

In order to prove rule (2's) properties, one has to define instrumentally an "energy function" (Hopfield 1982) :

$$E = -\frac{1}{2}\sum_{ij} w_{ij} s_i s_j$$

Intuitively, the function represents a measure of "tension" between the different hypotheses represented by the nodes' states. Pairs of nodes that are in the same state and are related by negative (inhibitory) connections will increase the state of cognitive tension within the network. Pairs of nodes that are in opposite states and are related by negative connections will clearly satisfy the inhibitory nature of their connection, thus relaxing the amount of tension within the network. An analogous consideration applies to the case of positive connections. In other words, the lower the energy in the network is, the higher the degree of "coherence" within the system of hypotheses will be, as represented by the current state of the network.

Because connections are symmetric, the energy function E can be rewritten as:

$$E = -\sum_{(ij)} w_{ij} s_i s_j$$

where (ij) stands for an unordered pair of i and j (in other words, 12=21).

Consider a new state of a node s_i:

$$s_i' = \text{sgn}(\sum_{j} w_{ij} s_j)$$

Clearly, whenever $s'_i = s_i$, nothing changes in the value of E. Thus, we have to analyze only what happens when $s'_i \neq s_i$. In that case:

$$E' - E = -\sum_j w_{ij} s'_i s_j + \sum_j w_{ij} s_i s_j$$

(we focus only on changes in E related to a single i)

Because in this case $s'_i = -s_i$

$$E' - E = -\sum_j w_{ij}(-s_i)s_j + \sum_j w_{ij} s_i s_j$$

$$= 2\sum_j w_{ij} s_i s_j$$

Extracting s_i:

$$= 2s_i \sum_j w_{ij} s_j$$

Since $s'_i = \text{sgn}(\sum_j w_{ij} s_j)$, whenever $s'_i \neq s_i$, either s_i is positive and $\sum_j w_{ij} s_j$ is negative, or vice versa,

Thus, whenever $s'_i \neq s_i$, E'-E is always negative because whenever $s'_i = s_i$, E'-E is always zero. Because **s** has a finite number of elements and the weigths w_{ij} are finite, E is bounded and we can thus conclude that the iteration of the update rule (2) always has a fixed point and will converge to it in a finite number of steps.