Organization Science
Articles in Advance, pp. 1–21
ISSN 1047-7039 (print) | ISSN 1526-5455 (online)

http://dx.doi.org/10.1287/orsc.2015.0987
© 2015 INFORMS

A Model of Collective Interpretation

Giovanni Gavetti
Tuck School of Business, Dartmouth College, Hanover, New Hampshire 03755, giovanni.gavetti@tuck.dartmouth.edu

Massimo Warglien
Department of Management, Ca' Foscari University of Venice, 30121 Venice, Italy, warglien@unive.it

We propose a cognitively plausible formal model of collective interpretation. The model represents how members of a collective interact to interpret their environment. Current theories of collective interpretation focus on how heedful communication among members of a collective (i.e., how much individuals pay attention to others' interpretations) improves interpretive performance; their general assumption is that heed tends to be uniformly beneficial. By unpacking the micromechanisms that underlie such performance, our model reveals a more complex story. Heedfulness can benefit interpretive performance. It can help collectives properly interpret situations that are especially ambiguous, unknown, or novel. Conversely, heedfulness also generates conformity pressures that induce agents to give too much weight to others' interpretations, even if erroneous, thereby potentially degrading interpretive performance. These two effects join into a nonmonotonic trajectory that represents how heed relates to interpretive performance: due to its beneficial properties, performance increases with heed until it peaks before degrading due to conformity pressures. The form of this nonmonotonic relationship is contingent on the nature of the task: ambiguous situations make collectives vulnerable to too much heed: ambiguity ignites conformism; novel situations make collectives dependent on heed: novelty requires multiple eyes to be seen. In addition to these results, our model offers a flexible platform that future work can use to explore collective interpretation in a variety of organizational and supraorganizational contexts.

Keywords: collective interpretation; neural networks; cognition; sense making; computer simulations; organizational memory
History: Published online in *Articles in Advance*.

Introduction

Interpretation is the cognitive act of giving meaning. It is thus essential to choice and action because what we decide and do cannot be decoupled from the meaning we give to the situations we are in. In organizations, interpretation is rarely the product of isolated agents. Whether we consider how organizational members interpret a competitor's move, customers' reactions to a new advertising campaign, an accident at a plant, or many other situations for which the meaning is unclear, interpretation generally reflects collective, distributed dynamics.

Yet, however much we know about interpretation at the individual level, we know much less about how it occurs in collectives. We know well-functioning collectives can achieve accurate interpretations in contexts that are too demanding for individual interpreters to manage independently, but what makes a collective function well is less understood. This gap is due in part to the phenomenon's complexity: the move from the individual to the collective introduces many extra variables that can affect interpretive outcomes, thereby posing challenges to theoretical tractability. Relatedly, although work on distributed processes has demonstrated the need to focus on the interrelationships among individuals, it has left individual cognition in the background, as if it were irrelevant to a theory of collective interpretation. This work has merit, but a theory of collective interpretation that abstracts from a cognitively accurate representation of individual interpreters is inherently incomplete.

A solution to this problem is to focus on a minimal set of variables that are foundational to collective interpretation, and that span the individual and collective levels. This article pursues that path by elaborating a formal model of collective interpretation, which offers a cognitively plausible characterization of individual interpreters who are engaged in collective interpretation. Through it, we perform a controlled analysis of some central causes of variation in interpretive outcomes. The model can also serve as a formal platform for future studies of collective interpretation.

Interpretation: Individual Level

Interpreting a certain reality means forming a representation of it—a conceptual structure in an individual's mind that encapsulates her simplified understanding of said reality (Lakoff 1987). This process is often associative, in that it is based on memory and perceived similarity (Lakoff 1987; Rosch 1978; Edelman 1987, 2006; Hofstadter 2001; Holyoak and Cheng 2011).[1] For instance, in the early 1940s, Charlie Merrill of Merrill Lynch reinterpreted the brokerage business using a supermarket lens (Perkins 1999). He formed a representation of brokerage premised on the perceived similarity between brokerage and supermarket, a business he knew well. This brokerage-as-supermarket analogy generated a new understanding of brokerage that in turn led to a path-breaking strategic innovation in banking (Gavetti and Menon 2015). In a typical associative process, the features of a new situation that the agent pays attention to evoke specific instances of past situations or higher-order concepts like categories (i.e., mental containers that cluster together similar things like objects, situations, or experiences) in her memory. In being evoked, these cognitive structures move from being "asleep in the recesses of long-term memory" to "gaily dancing on the mind's center stage" (Hofstadter 2001, p. 504), thereby becoming the basis of the new situation's representation. The defining feature of such processes is that they are based on similarity. Because of their cognitive efficiency relative to alternative interpretive mechanisms, Edelman (2006) argued that the brain evolved to make them hardwired in complex tasks that are central to human adaptation.

Indeed, theories of decisions that emphasize cognitive realism, whether in administrative disciplines, decision theory, economics, or political science, have increasingly recognized the importance of associative processes. For instance, in administrative studies, March (2006) suggests the need to move away from the imagery of rational choice, which portrays interpretation as a "model-based anticipation of consequences evaluated by prior preferences" (March 2006, p. 202), especially when we wish to understand how agents deal with situations that are novel and complex. In these contexts, agents can neither easily deduce models of the world, nor evaluate outcomes based on them. Similarity-based interpretation is more plausible (March 2006). Gavetti et al. (2005) share this premise in their analogy-based model of strategic decisions. In decision theory, Gilboa and Schmeidler (2000) challenge the anticipatory logic of expected utility theory on the grounds of its lack of cognitive realism: situations that are hard to interpret tend to be represented via past cases the decision-maker knows about. In economics, Mullainathan et al. (2008) model consumers as coarse thinkers who interpret new products in terms of broad categories. In political science, Neustadt and May (1986) argue that political leaders use analogies to think about domains that are uncertain and in which information is scarce.

Taken together, these contributions offer a strikingly convergent characterization of the cognitive bases of interpretation in worlds whose meanings are hard to construct. These models rest on the fiction of the lonely decision-maker. Therefore, their immediate usefulness to organizational questions is limited. Nevertheless, they offer a solid basis for building a microfounded model of collective interpretation such as the one we propose. Before articulating the premises of our model, we turn to collective interpretation.

Interpretation: Collective Level

Weick and Roberts (1993, p. 358) lamented that "[t]he preoccupation with individual cognition has left organizational theorists ill-equipped to do much more with the so-called cognitive revolution than apply it to organizational concerns, one brain at a time." Since this complaint, organizational scholars have been paying increasing attention to collective interpretation in organizations, where "collective" denotes a group of agents who are highly interdependent and interact frequently in their quest for meaning (Weick and Roberts 1993).[2] This field has been reviewed elsewhere (Walsh 1995, Meindl et al. 1996, Drazin et al. 1999), and we do not replicate such efforts here. We will, however, point out a few developments that are germane to our agenda.

The first development is research that documents the role that collective interpretation plays in many important processes in organizational life. This work has moved us from the *presumption* that collective interpretation is widespread in organizations because individuals tend to seek out others' interpretations when they face ambiguous situations (Goffman 1974, Volkema et al. 1996), to a more specific understanding of the situations in which it occurs in organizations. This work thus defines specific organizational situations that a model of collective interpretation can help us understand. For instance, collective interpretation or sensemaking has been shown to underpin the formation of strategies (Porac et al. 1989) and group decision making in top-management teams (Finkelstein et al. 2008, Smith et al. 2010). Similarly, it has been shown to be crucial to organizations' interpretation of focal issues in their environment (Dutton and Dukerich 1991). In the domain of innovation, a collective sensemaking lens has been evoked to explain failures and successes in product innovation (Dougherty 1992), or creativity more broadly (Ford and Gioia 1995, Drazin et al. 1999).

A striking illustration of the organizational relevance of collective interpretation can be found in work on high-reliability organizations (Hutchins 1991, Weick and Roberts 1993). This work shows that partially ignorant actors in a distributed system can accurately interpret complex situations when they interact appropriately. Stated differently, well-functioning collectives can be reliably effective in contexts that are so challenging that individual agents alone would likely make interpretive errors. For instance, when there are unexpected departures from routine operation in navigation or flight operations (e.g., a crisis), collectives have been found to properly recognize the new situation even when the available information is misleading or unreliable (Weick 1990, Hutchins 1995); or when critical pieces of information are missing (Weick 1990). Similarly, well-functioning collectives have been found to form accurate interpretations of situations that are fundamentally novel

(Michel 2007), or to absorb substantial turnover in the collective's composition (Michel 2007).

This type of processes are not exclusive of high-reliability organizations. Narduzzo et al. (2000) found similar principles at work in the emergence of troubleshooting routines in a cellular phone company. For instance, they showed that complex diagnostic tasks were interpreted more effectively when teams of troubleshooters developed routines for sharing their own individual representation of the problem to form a shared representation. In this vein, in their reconceptualization of routines as a source of both stability *and* change, Feldman and Pentland (2003) noted that collective interpretation is critical to organizational routines' ability to adjust to changing contexts, a position echoed by others (Edmondson et al. 2001, Obstfeld 2012, Cohen et al. 2014). In sum, this work suggests that collective interpretation is relevant to the microfoundation of organizational routines.

The second, complementary development is progress in the conceptualization of collective interpretation. This development has two components. The first involves a shift from the individual to *interactions* among members of a collective. This shift reflects the idea that in a distributed cognitive system, different members contribute distinct knowledge, and cognitive functions carried out by the system result from the patterns of interconnections among the system's elements (Weick and Roberts 1993). This theoretical conceptualization builds on work on transactive memory (Wegner et al. 1985, 1991), which emphasizes how collectives of agents can form generalizations of a complex phenomenon as a result of exchanges of disparate lower-order, detailed inputs. A related influence has been work in artificial intelligence on group mind (Sandelands and Stablein 1987, Hutchins 1990, Boden 1991, Rumelhart 1992), with its representation of collectives as distributed information-processing systems, and its emphasis on how the parts of the system (i.e., individual agents) interact.

The second component is the increased interest in explaining the drivers of interpretive outcomes, especially the accuracy and speed with which an interpretive system converges on an interpretation. In much of this work, variation in interpretive outcomes is explained in terms of variation in supraindividual variables that govern interaction (e.g., how agents interact, what culture is functional to collectives' proper functioning, etc.). Especially salient is the construct of "heedful interrelating," as emphasized by Weick and Roberts (1993) in their pioneering work on collective mind. In their study of aircraft carriers, Weick and Roberts (1993) argue that variations in heedful interrelating can generate profound variations in how well ambiguous situations are interpreted. Indeed, in most of the studies of high-reliability organizations that we mentioned above, the basic presumption is that high levels of heed are essential to make a collective effective in dealing with trying situations.

To sum up, these developments have (a) pinpointed organizational arenas in which collective interpretation plays a central role; (b) provided initial conceptualizations of collective interpretation as a process and as an outcome; and (c) demonstrated that understanding what makes a collective function properly can reveal first-order causes of variation in the quality of organizations' decisions and actions.

Toward a Model

We said at the outset that properly accounting for collective interpretation in organizations stretches the bounds of theoretical tractability. A collective interpretive outcome depends both on solo agents' cognition and on a process in which individual agents exchange information, influence others' interpretations, revise their own interpretations, etc. It is presumably because of this complexity that students of collective cognition have thus far not really expanded the primary level of analysis, but shifted it from the individual to the interrelationships among them. In doing so, they have de-emphasized individual-specific memories and the interpretive processes through which they are elicited and used. As argued by calls for increased attention to the individual's role in distributed systems (Resnick et al. 1997, Michel 2007), however, a theory of collective interpretation cannot be complete unless it rests on a cognitively realistic understanding of the individual. Any collective performance, Feldman and Pentland (2003, p. 109) remark, is "energized and guided by the subjective perception of the participants." We believe this gap is a key reason for the lack of a general understanding about what factors in how collectives function are particularly important to interpretive outcomes, and under what conditions of the interpretive task. For instance, we have just seen that heedfulness can help collectives achieve accurate interpretations in trying conditions, but it would be important to know if this is always the case, or if it can sometimes undermine interpretation. Similarly, it would be helpful to know if heedfulness is monotonically good, or if it can hurt beyond a certain level. More broadly, a key question is what role that variables beyond heedfulness (i.e., the homogeneity of the collective, the way it is structured, its power distribution, etc.) play in interpretation and under what contingencies of the task environment.

It is now possible to fill this gap because of the developments discussed in the prior section, which attenuate the theoretical tractability issue. We now have a robust yet parsimonious individual-level microfoundation of collective interpretation; initial conceptualizations of collective interpretation that can guide the development of a model of the collective; and empirical evidence, especially on the robustness of collective interpretation in trying situations, that can be used as a benchmark against which to test a model of the collective.

In this article, we propose a model of collective interpretation that spans a characterization of individuals who have a reality to interpret, and their relationships to other members of the collective in their quest for meaning. Our model tackles issues of theoretical tractability by seeking to focus on a minimal set of variables that are foundational to collective interpretation and interpretive outcomes. Its structure allows us to establish the precise role these variables play vis-à-vis interpretive outcomes.

With this premise in mind, we start by portraying the individual agent, and characterize the cognitive mechanisms involved in the recognition of the problem she faces. Our modeled agents' interpretation is governed by a memory that operates associatively over sets of features (Anderson and Bower 1980). That is, they encode the problem at hand in terms of qualitative features, and recognize it via an association to the past experience that most closely matches such features. To model these processes, we rely on the formal apparatus of associative neural networks (Hopfield 1982, Amit et al. 1994), which are thought to capture the basic properties of the neural processes involved in associative memory (see Miyashita 1988, Fuster 1995, Poucet and Save 2005, and Wills et al. 2005 for neurophysiological demonstrations of such properties and Amit et al. 1994 and McRae et al. 1997 for early attempts to reproduce experimental observations of human memory by associative neural network models). The success of such models in predicting individual behavior in multiple-cues decision making and probabilistic inference has been demonstrated by Glöckner and Betsch (2008) and Glöckner et al. (2010).

Second, we characterize interaction. Our model represents a collective of agents who share their interpretations of a reality with each other, and keep communicating until a stable collective interpretation is achieved. A process of this sort can be influenced by many variables, like the size of the collective, the heterogeneity of its members, its power structure, the organization of communication (i.e., who talks to whom), and variables of the reality to interpret, such as novelty and ambiguity. Our analytical structure is flexible enough to readily capture most of these variables. Here, we focus especially on the role of "heed" (i.e., how much A takes into account what B tells her) vis-à-vis the interpretive outcome. There are two reasons for our choice of perspective. First, heedfulness has a special status in both the theoretical and empirical literature on collective interpretation. Second, even before we use the model to generate novel predictions, we need to test its reliability against a benchmark of established empirical regularities: Does the model reliably generate predictions that square with empirical evidence? Work on high-reliability organizations provides such a benchmark.

Specifically, we first use this model to deduce, in closed form, general properties of collective interpretation. We demonstrate that heed significantly affects interpretive outcomes, and that within certain ranges, it leads agents to shared interpretations that are new to each of them and do not correspond to anything they individually know or have previously experienced. Collectives can be a source of genuinely novel knowledge. This property is important when agents face novel problems that do not match any of their previous experiences: heedful communication can enable a novel reality to be interpreted appropriately. However, heed can also have pathological effects. We demonstrate how excessive conformism may emerge when heed's intensity is too strong.

We then use the model to explore, via simulation, collective interpretation in less stylized settings. In particular, although we know that collectives can be effective even when information is noisy, or unavailable, or situations that are profoundly novel, etc., we know less about what makes a collective function well. By using our model to explore stressful settings, we obtain a fine-grained understanding of what can make collectives more or less effective in these situations. Expanding on the closed-form properties and allowing their quantification, our simulations suggest that the level of heed has a nonmonotonic effect on the quality of collective interpretation. Up to a certain level of heed, communication greatly helps collectives' effectiveness even in trying conditions. Beyond a certain level of heed, however, interpretive outcomes worsen as heed increases. Our simulations also suggest that the level of heed that brings a collective to maximum interpretive accuracy, and the sensitivity of a collective to the detrimental effects of excessive heed are contingent on the nature of the interpretive task. Novel situations require higher levels of heed for novelty to be recognized as such, and peak performance is thus achieved for higher levels of heed. Conversely, ambiguous situations offer agents many opportunities to form erroneous perceptions of the situation, which make them especially vulnerable to too much heed. Furthermore, simulations allow to explore the process of convergence to an interpretation, further illuminating how different conditions affect the final outcome.

The article's contribution involves both fostering knowledge of collective interpretation and offering a formal platform that future research on organizational interpretation can use to explore a wide variety of situations. The merit of the first objective is self-explanatory. Let us spend a few words on the second objective. Although a set of core variables can be thought of as being invariantly important to interpretation in most settings, other factors can play roles that vary widely across situations. For instance, a collective of top managers seeking to make sense of an ambiguous strategic problem faces a different problem than a crew of firefighters facing a crisis does. In the former case, the composition of the collective (e.g., the diversity among members' histories, backgrounds, memories) can significantly influence how the collective makes sense of the problem (Gavetti et al. 2005); in the latter case, the interpretive outcome might hinge more critically on how cognitive labor is divided among firefighters (i.e., what facets of the environment each firefighter is asked to pay attention to), and the timeliness with which the information gathered is shared among crew members (Weick 1993). Although we look at collective interpretation through a particular lens, we offer an analytical structure that can be used to represent different situations because collective interpretation can take different forms. (Online Appendix 1, available as supplemental material at http://dx.doi.org/10.1287/orsc.2015.0987, reports a few analyses reflecting configurations of collective systems other than ones we considered in the main text of the paper. It thus gives a flavor of the versatility of the model.)

Premises, Assumptions, and Modeling Choices

Premises and Assumptions

Associative interpretation is the representation of a situation based on its perceived similarity with something the agent already knows or has experienced (i.e., a cognitive representation of a situation or experience that is stored in her memory) (Anderson and Bower 1980). Therefore, any cognitively plausible model of associative interpretation needs to capture realistically three key elements of this process. First, *cognitive representations*, which are the foundation on which associative interpretation is based—both the raw materials used in the interpretive process and the outcomes derived. Second, how individuals *perceive similarity* between what they know and the new reality they need to interpret. Third, *memory* operation, especially how experiences are stored and how they are accessed in interpreting reality. Moving from the individual to the collective level then requires that we address how *communication* among individual interpreters generates meaning. The theoretical premises and assumptions that we make in dealing with these challenges are summarized below.

Cognitive Representations. Cognitive representations are conceptual structures in individuals' minds that encapsulate a simplified understanding of the reality these individuals face (Lakoff 1987, Thagard 2014). More concretely, a long tradition in cognitive psychology (Tversky 1977, Rumelhart and Ortony 1977, Gentner 1983) treats cognitive representations as clusters of features. That is, out of the many possible features along which a given reality can be characterized, a cognitive representation is a low-dimensional set of such features—a coarse characterization of an infinitely detailed reality.

Perceived Similarity. Associative processes are based on perceived similarity between different realities (see Larkey and Markmann 2005 for a broad discussion). In cognitive terms, this process consists of drawing a link between the reality to be interpreted as captured by an initial, embryonic representation (i.e., the subset of situational features the agent observes or infers) and a more fully formed representation of a prior reality the agent has already experienced or learned about in some form. This link thereby connects cognitive entities, specifically cognitive representations. Similarity has been argued to be typically feature-based (Tversky 1977).[3]

Memory. The individual's memory is a collection of representations of situations that are accessed or retrieved in a direct way by associative processes (as in content addressable) and not by serial, systematic search over all memory addresses. Thus, memory can be characterized as a combination of a state (what is remembered) and a process (how memory states can be retrieved). In a typical associative process, an agent is presented with a situation to interpret, and some observed or inferred features of the situation evoke a memory state in her mind. Neurobiological studies of memory (Kandel et al. 2000) offer strong empirical support for the idea that memory retrieval proceeds associatively (Anderson and Bower 1980, Edelman 1987). Specifically, associative memory is a content-addressable process that directly retrieves a memory that closely matches the reality to be interpreted along some features observed by the agent (Anderson and Bower 1980). A classic example is how a flavor can instantly evoke entire episodes from the past. Such a connection is relatively immediate and does not require a systematic search through a bank of memories. This characterization of memory operation improves on what is commonly assumed by models of analogy-based interpretation, which suggest some kind of exhaustive, brute force serial search through all possible direct and vicarious experiences stored in the agent's memory. For example, case-based decision theory (Gilboa and Schmeidler 2001) assumes that agents look through *all* cases stored in their memory, and the Gavetti et al. (2005) computational model adopts a similar perspective.

So far, we have discussed interpretation in terms of associations with memories of single experiences. Interpretation often proceeds, however, not on the basis of associations with individual experiences, but in relation to categories, i.e., mental structures that group anything experienced or learned along some dimensions of similarity (Rosch 1978, Lakoff 1987). Categorization—interpretation of a new reality by its association with a category—is cognitively analogous to association with singular experiences. When agents categorize a situation, they associate it with a prototypical exemplar of a certain category. That is, the similarity between the situation to be interpreted and a given category is assessed with reference to a specific prototype of the category. Once the situation is categorized, it is represented analogously to the category's other members. Thus, associative interpretation can operate in relation to both individual experiences and categories. Our model aspires to be general enough to capture both forms.

To sum up, in our characterization of interpretation at the individual level, we assume that; (i) cognitive representations are feature-based; (ii) associations are based on similarity, which, in turn, is feature-based; and (iii) memory is a collection of representations that are accessed or retrieved associatively and not serially.

From the Individual to the Collective. To move from individual to collective interpretation, we need to account both for communication and for how the sharing of individual interpretations via communication translates into a collective interpretation. Consistent with the assumption that representations are feature-based, we assume communication operates over features. Individuals tell each other what features of the reality they "see," and they can be more or less heedful to what other individuals in the collective tell them. The collective achieves an interpretation when none of its members feels compelled to challenge others' opinions.

In addition to these two basic elements, we need a structure that allows us to capture some of the variables that reflect the possible contingencies that can affect interpretive outcomes. Agents can be more or less homogeneous in their cognition and experiences; they can be more or less focused in the attention they pay to external cues; the structure of communication can be more or less centralized, or more or less hierarchical, etc. Information from the environment can be more or less noisy, more or less complete, more or less novel, etc.

Modeling Choices

Neural networks techniques capture both associative processes and feature-based conceptions of representations and memory (Hertz et al. 1991). These techniques are empirically robust: when used to reproduce experimental observations on human memory, they consistently demonstrate strong explanatory power (Poucet and Save 2005).[4] Among the various neural network approaches, the Hopfield model (Hopfield 1982) is the simplest model of content-addressable memory retrieval. As we explain in detail below, the Hopfield model represents both novel situations to be interpreted and experiences stored in the agent's memory as clusters of correlated features. The memorized situation most similar to the input stimuli (the features of the novel situation that the agent pays attention to) is retrieved through a process that exploits correlations among features to converge to a memorized situation without visiting the agent's full memory. Because this model is empirically robust, mathematically tractable, and consistent with our assumptions, we build on it to model individual interpretation.[5] Nonetheless, all modeling choices entail crucial assumptions and simplifications. Some of the choices we made imply minor losses of generality. For example, continuous or stochastic activation rules can be adopted instead of the threshold function we adopt (see (1) below) that preserve the fundamental associative memory properties of the model (Hertz et al. 1991). More significantly, introducing strongly asymmetric connections might affect the stability of the network, but offer opportunities to memorize sequences and temporal patterns. Consistent with the "minimalist" nature of our modeling effort, we have opted for the simpler model. Although we are aware of its limitations, we decided to trade off the formal complexities of more elaborate models because we believe that such simplifications are justified by their heuristic fruitfulness (see however the Online supplement for the exploration of some more complex model).

Neural networks also naturally lend themselves to representing the relational component of interpretation, as the analytical intuitions of the connectionist school suggest (see especially Hutchins 1995). We build on such intuitions and construct networks of neural networks—collectives of agents who face a new situation that is initially interpreted on an individual basis before agents communicate what they see to each other until a stable state—a collective interpretation—is achieved.

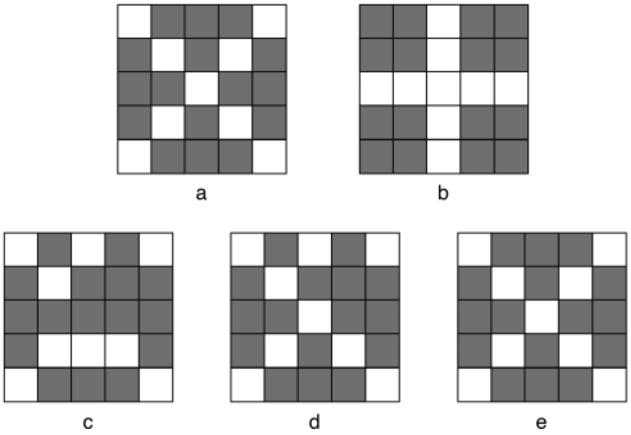
The Model

We first lay out the model of the individual, and then use it as the microfoundation for the collective-level model.

The Basic Setup: Individual Level

Intuition. Before articulating our formal model of individual-level interpretation, we preview its central mechanisms informally. The agent is given a reality or situation to interpret that she represents or recognizes associatively: The situation is recognized when a memory in her mind is activated. This memory will be the basis for her interpretation. In the model, the reality is characterized as a set of features, and so is the agent's memory, which consists of a set of prototypical experiences, each of which is encoded via features that might or might not be present. For instance, if the agent is a strategic decision-maker operating in a new business, she will interpret this setting in terms of her memory of prior businesses. She will observe some features of the present business, and activate a memory of a business that is similar along the features she observes. The agent's experiences are stored in her memory as a neural network, with each node of the network representing a feature, and the connections among them encapsulating the agent's experiences. 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

Figure 1 Associative Recall: A Visual Example



present across the agent's experiences, their connection will be heavier than it would be if the two features did not co-occur. This fundamental principle is known in neuroscience as Hebb's Rule (Brown et al. 1990). With this setup in mind, it might be useful to regard the nodes of the network as hypotheses about the presence or absence of features. When the agent has a new situation to interpret, the network will be initialized to reflect the new situation (i.e., each node will reflect hypotheses about the presence or absence of features according to what the agent initially sees). 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 making and the agent's memory. This iterative process will continue until the network converges to one of the experiences that the agent stores in her memory.

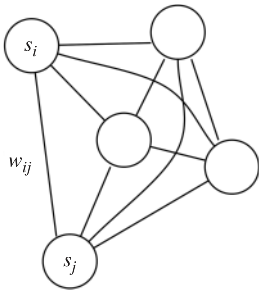
A simple visual example will help clarify the nature of the associative interpretation process. Consider a memory that stores two prototypical patterns, as in Figures 1(a) and 1(b).

Each configuration can be represented as a vector of 25 nodes (the little squares). Once the memory is stimulated by a new visual input (Figure 1(c)), 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 1(e)).

The model has three building blocks: environmental states (i.e., realities to be interpreted), which we call *situations*; agents' memories (i.e., limited repertories of situations); and network dynamics (i.e., the process through which the neural network associates a given situation with a given memory). We describe them in turn.

Situations. We model situations or realities as configurations of features, and assume that the set of such features, $F = \{f_1, f_2, \ldots, f_n\}$, 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. There are $Z = 2^N$ conceivable situations.

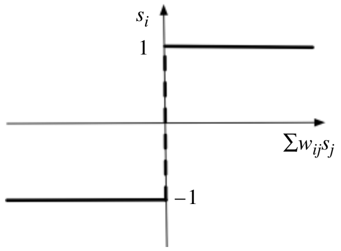
Figure 2 A Neural Network



Memory. Individuals' limited memory comprises a repertoire of situations (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: $M \subset Z$ and $\#(M) \ll \#(Z)$. We model an individual memory as a neural network, which is made of nodes (i.e., artificial neurons) that can fire or become active when incoming stimuli exceed some threshold. Consistent with the interpretation we suggested above, a given node fires when the hypothesis is accepted that the feature associated with it exists. Nodes are connected by arcs (i.e., artificial synapses) 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 translated into a pair $(\mathbf{s}, \mathbf{W})$ where $\mathbf{s}$ represents the nodes' states, and $\mathbf{W}$, the matrix of weights, represents the adjacency network of the network graph. Formally, $\mathbf{s}$ is a vector of N binary variables s_i that take on values $\{1, -1\}$, and $\mathbf{W}$ is a symmetric $N \times N$ matrix of real-valued weighted connections w_{ij}.

Network Dynamics. Given a matrix of connection weights $\mathbf{W}$, 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 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 $j \neq i$, weighted by the strength of the connections from the jth node to the ith 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 (or 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

Figure 3 Update Rule



to take the sign of the aggregate input to determine the value of the *i*th node (see also Figure 3):[6]

$$s_i = \mathrm{sgn}\left(\sum_j w_{ij} s_j\right). \tag{1}$$

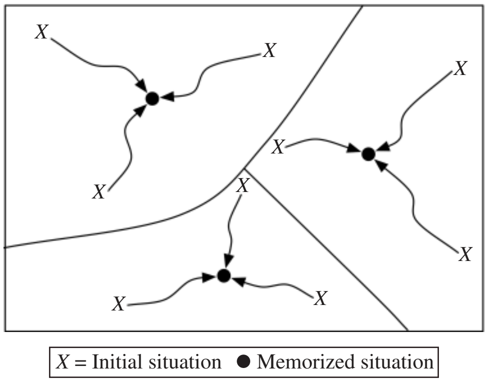
Rule (1) has two useful implications:

(i) It has been proved that it always leads the network to a configuration in which no node state is changed or modified: a fixed point. Because of this property, the agent's memory can be represented as a set of situations stored as fixed points.

(ii) 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 (technically speaking, to the fixed point with the lowest Hamming, or bit by bit, 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 shows an idealized two-dimensional representation of such decomposition in basins of attractions around the stored situations.[7]

Because a stored situation will be always recalled at the end of an interpretation process when the new input is close enough to it (i.e., it falls into its basins of attraction), associative memories have two important properties. First, they have an error-correction property: If a situation is distorted by noise or errors when it is presented to an agent, the memory will recall the "right" situation (as in the visual example of Figure 1). Of course, if the error is so pervasive that it makes the input look more similar to another situation stored in memory, the memory will misinterpret the stimulus and retrieve the latter situation. Second, for the same reason, associative memories can fill information gaps such as incomplete inputs (i.e., features that are initially unspecified).

Figure 4 Basins of Attraction



Note. Adapted from Hertz et al. (1991).

Storing Situations. Although nodes update their state during the recall process, we keep **W** constant. Thus, the connection weights are the parameters of the network dynamics. This condition implies that a situation can be memorized by appropriately tuning **W**. In the language of neural networks, this memorization 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 (Hertz et al. 1991)). Alternatively, we can directly engineer the storage of situations as fixed points in agents' memory. Because our focus in this paper is not on agents' memorization processes, but rather on how a *given* memory is used to interpret new situations, we adopt the latter approach.[8]

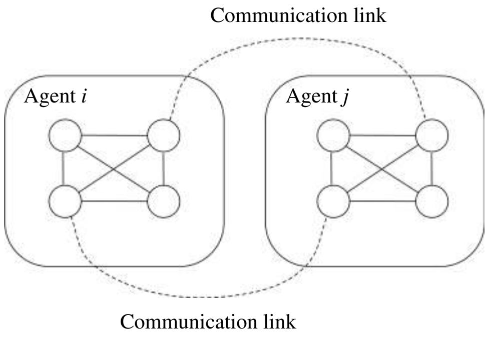
From the Individual to the Group

Core Structure. We characterize the group as a set of individual interpreters with a communication structure among them. In a nutshell, the group faces a situation to interpret. Each agent in the group first interprets the situation individually. Agents then share their initial interpretations with other members of the collective, and initiate a process of mutual influence until the group achieves a stable collective interpretation.

We model communication through the architecture introduced by Hutchins (1995) and further developed by Marchiori and Warglien (2005). In this architecture, every agent receives signals from others about their current interpretation of the environment (as represented by the current configuration of their own memory network). Signals focus on specific features (nodes). For instance, if agents 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 message adds to the input received by agent 2 on f_i. For example, suppose a U.S. company is considering entering a Latin American region, which is new to this company and its competitors. Suppose senior managers are evaluating the situation collectively. The key challenge is to interpret what the region will be like (e.g., what type of customers, what institutional environment, what infrastructures, etc). Our model would capture this situation by having each executive form a first impression of the situation based on her own perceptions. Then, in some asynchronous order, each executive would start re-evaluating single features in light of their coherence with the patterns ("cases") stored in her memory, but also taking into account what she knows about others' perceptions. Each executive would then tell her peers what she believes is going on, as she updates her opinion of the focal feature. More specifically, let us consider how a typical executive in our model communicates with her peers (the same will happen to all executives in the group). This executive will first try to get a sense of the current beliefs of her peers on a given feature (e.g., the type of consumer prevailing in the region). She will compare these with her own perception and the overall patterns stored in her memory. She will then update her beliefs about such features while trying to find the highest coherence between her own perception, her memory, and the beliefs of her peers. She might thus reject some suggestions outright because they are deeply inconsistent with her memories; but others might cause her to change her opinion for some specific characteristics. For example, she might find that given the type of mindset that she believes to be dominant in the country, the type of institutional environment the firm will likely face, the other features she considers, and the opinions of the majority of other managers, her current perception about the type of consumer might be misleading. She will thus update her beliefs about that feature. She will then communicate to her peers her new view about that feature. This process will continue iteratively, with managers asynchronously updating their beliefs about specific features, until a stable interpretation is achieved. This process of communication is represented by connections between nodes of the agents' respective memories, which transmit signals from one agent to another. A given agent will simply add the weighted signal from another agent to those internally generated by her own memory when she updates her interpretation of a feature. Figure 5 shows a two-agent communication structure (dotted lines), in which agents communicate over two features.

To concretely affect interpretation, communication has to be *intense* enough to affect the state of mind of the recipient of the message. An increase in the intensity of communication means that the recipient of a message pays more attention to the message and is more affected by it. In our model, this effect is modulated by the weight of the connection between agents: The heavier the connection, the stronger the message's effect on the recipient's update of her interpretation. We call this effect "level of heedful communication" or in short "level of heed."[9]

Figure 5 Introducing Multiple Agents



Moving to the formal structure, the group-level model could be described as a network of neural networks. The whole group itself is a large collective 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 between-agent connections or level of heed. In the analyses that follow, we assume that γ is the same for all between-agent connections. In concrete terms, 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 expressing how much group members "pay attention to each other." By tuning γ, we can therefore obtain groups characterized by various degrees of internal influence, which correspond to various degrees of collective heed, thus summarizing the potentially infinite and diverse factors that determine the influence of communication among group members.

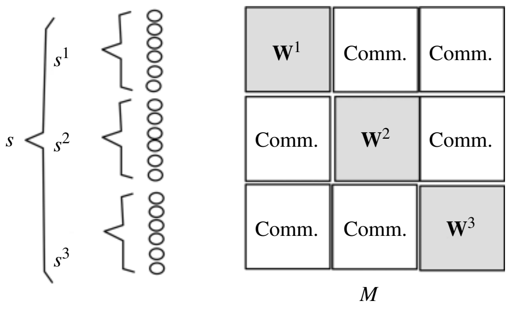
The matrix $\mathbf{W}$ of all such connections is a symmetric square matrix that has the (shaded) blocks constituted by each agent's internal connections $\mathbf{W}^k$ (i.e., individual memories) on its main diagonal, and the blocks representing communication among agents outside the main diagonal (see Figure 6).

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

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

A Flexible Structure. This analytical engine offers a flexible platform that can be used to characterize a variety of collectives facing a variety of problems. For instance, γ can be easily tuned to explore the effects of asymmetries in intragroup heedfulness, such as those

Figure 6 A Multiagent Neural Network



that reflect some form of leadership. Furthermore, communication can be more or less *dense* as dictated by which agent talks to which other agent within a group. Also, communication can differ in *scope*. That is, a pair of agents can communicate over a more or less extensive set of features. These two parameters, density and scope of communication, allow us to model different communication structures. For instance, one can represent 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 with only a central agent, or many other structures corresponding to different communication patterns. 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. Indeed, it is possible to model individuals that are heterogeneous in their expertise by simply controlling how homogeneous the experiences they have in memory are. We can also control the type of environment the agents face. Some environments are noisier than others are. Information can be corrupted, or it can be incomplete, etc. Given our goals for this article, we explore only a subset of these contingencies. However, to provide a flavor of the versatility of the model, Online Appendix 1 summarizes a few additional analyses that reflect configurations other than the ones we consider below. Most notably, we study how collectives with strong leadership behave vis-à-vis structures of peer agents.[10]

Analysis

We use this formal structure to first derive, in closed form, some fundamental properties of collective interpretation that we expect to be robust across contexts. We derive these properties by considering the simplest parametric configuration of our model: full communication among agents and a noiseless environment. We then turn our attention to more realistic settings, which we analyze via simulation.

Some Fundamental Properties of Collective Interpretation

To start, we need to define collective interpretation precisely. By collective interpretation, we mean a stable state in which no agent has reason to modify individually her current interpretation of the environment. Analytically, we model collective interpretations as fixed points of an effort in which individuals modify their own interpretations to satisfy the pressures of both their own memories and the opinion of others. Note that this definition does not require that all agents share the same interpretation—they might agree to disagree: there may be fixed points of the group network in which agents have persistently 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 associative pressures coming from her own internal mental states with those from others' interpretations—a kind of collective reflective equilibrium (Goodman 1955, Rawls 1971). The case in which all agents reach the same interpretation, which we label *shared interpretation*, is a special type of collective interpretation, and will play a major role in our analysis. Notice also that this definition makes an implicit but clear distinction between interpretation as a state and the process through which this state is reached (a same state might be reached through different processes).

Any model of collective interpretation should address a few fundamental questions. The first, most obvious question corresponds to the *existence problem*: Can collective interpretation as we define it be achieved in our model of the collective? Without a positive answer, our modeling effort would be meaningless. The second question corresponds to the *consensus problem*: Can shared interpretation be achieved, and, if so, what factors enable shared collective interpretations? The third question corresponds to the *creativity problem*: Does collective interpretation have to reflect a pre-existing interpretation of at least one agent in the group, or can genuinely new interpretations emerge from communication among agents? If new interpretations can emerge that are shared, it would be important to establish under what conditions collectives can lead to creative insight. Finally, a tradition in the study of groups suggests that there may be negative side effects of collective interpretive efforts, a tendency of collectives to generate conformity—a tendency to unanimity that overrides the goal of realistically appraising situations (Asch 1957, Janis 1972, Baron 2005). A final question thus concerns the *conformity problem*: can the model express such collective conformism property and, if so, under what conditions? Below, we provide a general, closed form answer to these questions in form of four propositions. Proofs of the propositions can be found in Online Appendix 5.

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 interpretation of the environmental input). The answer is positive, and we express it as the following:

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

We offer an intuitive explanation that is based on Figure 6 and expression (2): The group model has the same formal structure as a Hopfield network, with an update rule that includes other-agents' inputs. The group is thus 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 properties of the individual model, especially the existence and local stability of fixed points, which act as group memory states.[11] 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 interpretation.

Propositions 1 provides no information about the possibility of shared interpretations or what conditions are necessary for their existence. Intuitively, as agents pay more attention to others' interpretations, the pressure to agree on the same interpretation increases. We can prove an even stronger statement than this intuition: If the level of heed is high enough, consensus will always emerge. Accordingly, Proposition 2 states the following:

Proposition 2 (Consensus). *Provided that γ is large enough, ALL collective interpretations must be shared ones.*[12]

The reader is referred to Online Appendix 5 for a proof of Proposition 2. The intuition of the proof is that when a group is in a very high γ condition, the majority opinion on each feature will spread in the group, forcing an agreement.

Propositions 2 opens the door to the third question: If communication can force consensus among group members, is this effect limited to inducing agreement only on pre-existing interpretations, or can genuinely novel interpretations instead emerge out of agents' communication? We can prove the following (see Online Appendix 5 for the proof):

Proposition 3 (Creativity). *If γ is large enough, there will always be a shared interpretation that does not correspond to any of the situations stored in individual memories.*

Proposition 3 indicates a genuinely creative process for generating new interpretations of the environment. Heed can induce 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 when the situation to interpret is truly different from anything experienced before—a whole new truth is created out of partial ones. Ironically, here recognition (seeing the new in terms of the familiar) may lead to new cognitions.

The dynamics that underlie Propositions 2 and 3 suggest that groups characterized by very high levels of heed may be induced to quickly lock into initial interpretations, no matter how arbitrary they are. As γ increases, they may become increasingly "credulous," prone to agree on everything—a state reminiscent of the pathologies of pressure to conform. Consensus may override realism. In fact, we can prove the following:

Proposition 4 (Credulity). *Provided that γ is large enough, ANY arbitrary shared interpretation can be a fixed point.*

In sum, Propositions 2–4 suggest that an increase in collectives' heed has both positive and negative effects on interpretive outcomes. On the positive side, Proposition 3 guarantees that genuinely new situations can be identified as such. If no agent has memories that closely correspond to the new situation to interpret, intense communication may help generate an interpretation that is closer to reality, and it does so by compensating for individual errors through the emergence of collective wisdom resulting from aggregating right hypotheses that are diffused among different agents. Thus, communication can reinforce the error-correction and information-filling virtues of individual associative memories (see also the next section). On the negative side, as γ moves beyond a critical threshold, the pressure to conform can lead the group to converge to any arbitrary configuration.

Although these propositions establish fundamental properties of collective interpretation, they also leave some questions open. To what extent can heed make collective interpretation more reliable by improving the error-correction and information-filling virtues of individual associative memories? Can it discriminate between truly novel situations and accidental perturbations of old ones, or will the creativity property result in mere noise filling and systematic misinterpretation? To what extent can conformism override the benefits of consensus? The next session will use computer simulations to consider a set of more specific and realistic task environments in which such questions can be investigated, and that also relate to relevant debates in the organizational literature.

Simulation Experiments: Collective Interpretation and Organizational Reliability

We explained at the outset that properly functioning collectives have the potential for reliable interpretive performance even in extremely challenging settings. Yet our current understanding of how such outcomes come about is limited. Work on collective minds (Sandelands and Stablein 1987, Weick and Roberts 1993, Weick et al. 1999) has pointed to the importance of some properties we discussed in the prior section, such as how heedful communication among agents can reinforce the properties of error correction and information filling of individual associative memories, favor the integration of conflicting information and representations, and generate new interpretations that do not already exist in agents' individual minds. This is a promising convergence, and we use our model to explore these properties more fully in four settings that are prototypical to the literature on organizational reliability and distributed systems, settings that induce misinterpretations if not properly handled.

The first setting is one in which the information upon which interpretation is based is noisy, in that it is corrupted and potentially misleading (Hutchins 1995). For instance, a cause of the Tenerife airplane crash was the Pan Am pilot's difficulty in comprehending what a "ground controller who spoke with a heavy accent" was trying to communicate (Weick 1990, p. 130). The second setting corresponds to situations in which critical pieces of information are hidden to the interpreter. To stay with the Tenerife accident example, another cause of this crash was the presence of thick clouds and fog, which made it impossible for the pilots to obtain critical information. The third setting is one in which, independent of the quality and amount of information processed, the situation to interpret is new to the interpreters. That is, it might not correspond to any of the agents' experiences or memories (Michel 2007). The final setting corresponds to situations in which a distributed system is disrupted by turnover in its members,[13] which can challenge interpretation when the incoming agents share little in terms of background, experiences, and beliefs with the other group members (Weick and Roberts 1993, Michel 2007).[14]

Basic Architecture

Computer simulations use the formal structure detailed in the model section, which has three building blocks: an environment of situations to interpret, agents, and a communication structure among them. We now briefly describe how we operationalize them, leaving a more detailed specification to the simulation description.

Environment. The environment is represented as a set of situations to interpret, each operationalized by 12 binary features.[15] We consider 3 situations (that represent true states of the world) that equally span the state space in terms of Hamming distance. The agents receive signals from the environment in the form of 12 bit strings, where each bit represents a feature. See below how we operationalize noisy and incomplete signals.

Agents. Each agent is modeled as an associative memory that stores a limited number of prototypical situations (in our model, the agent stores 3 situations), each consisting of 12 features. A prototypical situation is a 12-bit array stored as a fixpoint of the individual memory. Agents may store partial knowledge of the situation (not all features may be represented in their memory), or may have memorized situations that differ from individual to individual.

Group Structure. We consider a group of four agents fully communicating with each other. As specified in the Model section, communication occurs through feature-by-feature, weighted connections between agents. Groups' level of heed is modeled through γ. If γ takes on a 0 value, we speak, for simplicity, of no communication. Through this basic structure, we consider four settings:

(I) *Noisy Environments*. The situations to interpret can be noisy or randomly corrupted. We capture noise by randomly flipping individual "true" features, with noise being independent for each agent (i.e., each agent receives independently corrupted information). A *noise* parameter η determines the probability of a random flip of a feature as perceived by agents, and therefore the amount of noise that exists in the system.

(II) *Incomplete Information*. Agents can receive incomplete information about the situations to interpret. We capture this effect by assigning some features an uninformative 0 value. We model incompleteness as being independent among agents (i.e., agents receive independently incomplete information). We assume independence because it allows drawing conclusions that are not conditional on specific patterns of incompleteness. An *incomplete information* parameter i determines the probability that a given feature is uninformative.

(III) *Novelty*. The situation to interpret can be more or less novel. We model the degree of novelty in terms of the Hamming distance between true state of the world and prototypes stored in agents' memory. The more the true state of the world is Hamming distant from the closest stored prototype, the higher its novelty.

(IV) *Turnover*. Collectives' membership is not always stable. An important question is therefore how personnel turnover affects interpretive outcomes. If the incoming and outgoing members share the exact same experiences, the effects of turnover should be negligible. An interesting case of turnover is therefore the case of heterogeneous newcomers. Heterogeneity can be operationalized by endowing agents with different memorized prototypes. Accordingly, in the turnover treatment, three agents (the incumbents) have homogeneous memories, while the newcomer is modeled as having entirely different prototypes in memory.

Performance. In what follows, we take the rate of correct shared interpretations in a given environment as the main indicator of performance, measured over a large number of simulation runs (although we also consider other performance measures). This emphasizes the fact that in the prototypical environments that we explore, agents essentially play a coordination game, where success depends both on accuracy and consensus, and common interest is assumed. However, the simulation platform might easily be accommodated to different payoff structures, where for example the degree of consensus or the average accuracy provide additional sources of reward (and incentive misalignment among agents might be represented).

Simulation Results

I. *Noisy Environments.* Figure 7(a) summarizes the main results of simulations experiments under condition I. The simulation we conducted (Figure 7(a)) represents all agents as receiving complete, albeit noisy, information. We ran 10,000 iterations for each parameter combination of noise level η and communication γ (5×26); η takes values from 0 (no noise) to 0.2 (20% of chance of distortion for each feature) with 5% steps, and γ ranges from 0 (no communication) to 2.5.[16]

As can be seen, the interpretive performance is always (and trivially) correct when there is no noise because all agents receive the same signal from the environment, which matches a prototype that each agent has memorized. Heedful communication has little to add.

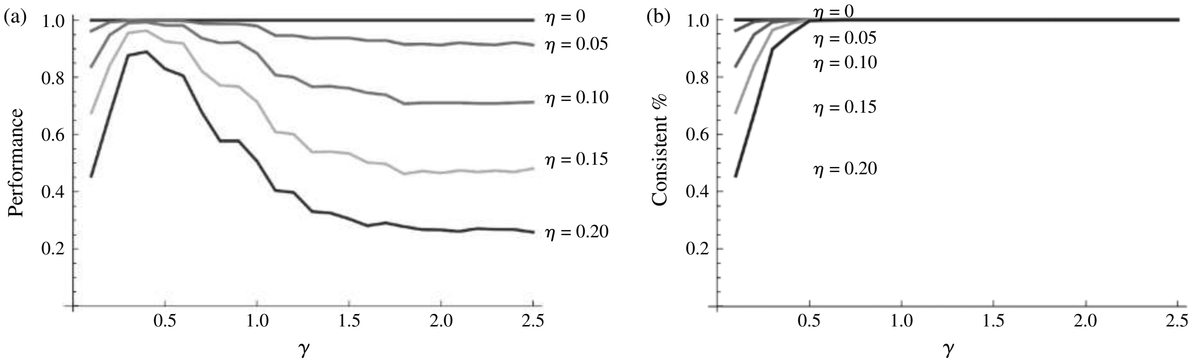
Noise has a significant impact on interpretive performance. Without communication, agents converge to an already memorized prototype, but often to the wrong one. Such an impact is small for low levels of noise (i.e., $\eta = 0.05$) because each agent has enough redundancy built in her memory to correct minor deviations form the right prototype. As noise increases, performance deteriorates (more than proportionally) because individual agents are misled by noisy information and may therefore retrieve an erroneous prototype from their memory.[17]

Thus, group failures in the absence of communication result because the error-correcting properties of individual associative memories are insufficient. On this ground, heedful communication has a powerful impact on interpretive performance. As can be seen from Figure 7(a), heed enables agents to reach almost error-free performance even for rather high levels of noise. The basic effect of communication among heedful agents is to reinforce the error correction and pattern reconstruction properties of individual memories. This is obtained by exploiting the redundancy (a constant trait of high reliability organizations) built in the collective. If an agent perceives a feature incorrectly, her misperception is likely to be corrected by agents that perceive it correctly. This is a clear reflection of the "consensus" property of heed analyzed in the closed form section. As long as the majority of the agents are right, there will always be a level of heed that makes the whole team right.

However, this beneficial effect exhibits an inverted-U-shape behavior: It declines after a critical level is reached. This decline is not due to miscoordination of individual interpretations among agents, as happened in the no- (or low-heed) communication case. After the maximum level of performance is reached, there is always final consensus (Figure 7(b)). What happens instead is that agents converge on arbitrary interpretations that "fit" noise. If noise is high enough, too much heed can be worse than no heed. This result is clearly related to the "credulity" property analyzed in closed form in the prior section.

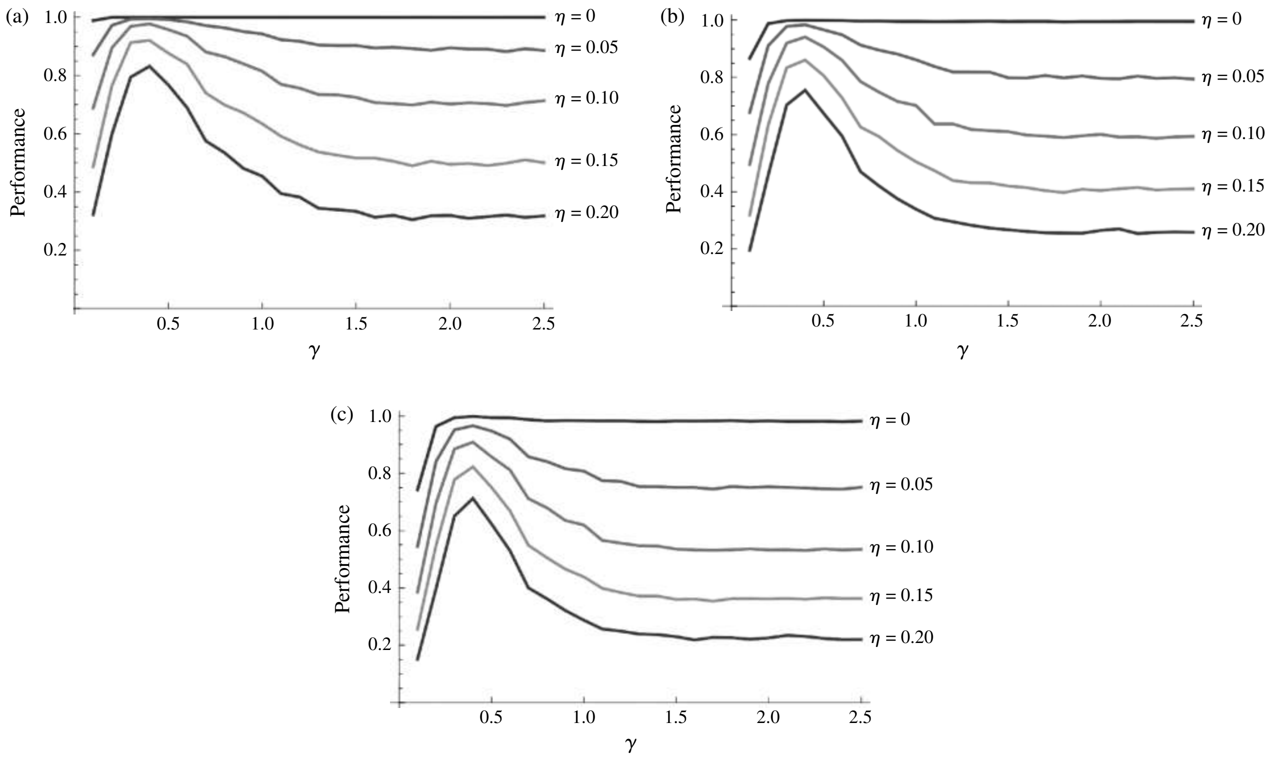
Thus, taken together, Propositions 2 and 4 result, within the environment that we have designed here, in a parabolic effect where the consensus benefits are overridden by the damaging effects of pressure to conform.

Figure 7 Noisy Environments



Note. Average performance (rate of correct interpretation: 1.0 = 100% of successes) and rate of convergence to interpretations consistent among agents (rate of consistency: 1.0 = 100% consistent interpretations).

Figure 8 Incomplete Information



II. *Incomplete Information.* We can now introduce incomplete information (Figure 8). We ran 10,000 iterations for each parameter combination of noise η and heed γ under different assumptions about incompleteness, the probability of a given feature to be uninformative being, respectively, 0.2, 0.4, and 0.6 (Figures 8(a), 8(b), and 8(c)).

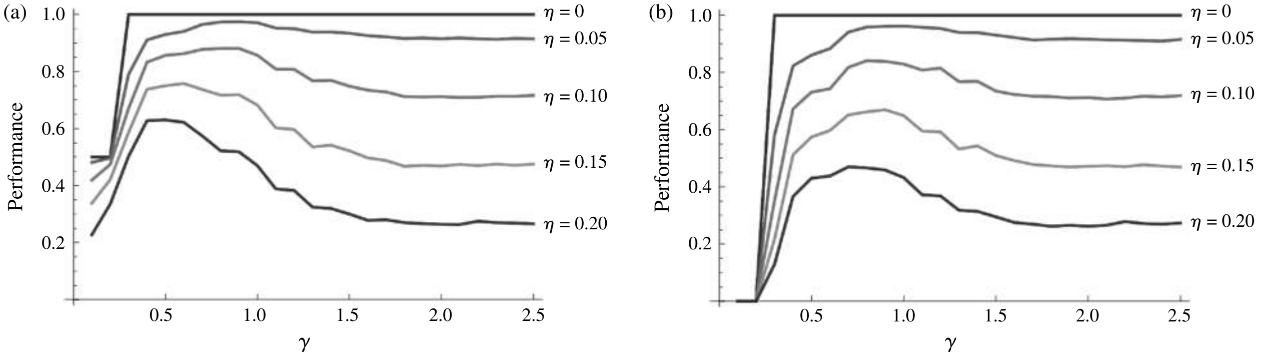
Figure 8 suggests a few relevant properties. First, incompleteness by itself has a much milder effect than noise does, as suggested by the comparison of the effects of noise and incompleteness (in the absence of noise) for the same probability level. For a noise probability of 0.20, without communication, the interpretive performance degrades below 50%. For the same probability of incompleteness, performance stays close to 100%. Only high levels of incompleteness generate substantial degradation of performance without communication. Furthermore, although noise tends to generate nonmonotonic effects of communication, heed is always beneficial when there is only incompleteness because incompleteness creates missing information, whereas noise creates false information. Thus, incompleteness does not have the distracting properties of noise. It can damage individual interpretation only when it creates enough information gaps that there are unsolvable ambiguities regarding which situation the available information may represent. Furthermore, when there is communication, it can damage collective interpretation only when features are missing for all agents, and no one agent can help the others to fill the gaps.

The interaction of incomplete information and noise is interesting. The more information for some features is missing, the less right information can correct distorted information, thus amplifying the effects of noise. Consequently, when communication is absent, information incompleteness further deteriorates interpretive performance. Nonetheless, incompleteness also dampens the credulity effect: For a given level of noise, it reduces the absolute number of wrong features (because information gaps cannot be turned into false information by noise) and thus reduces the potential for generating credulous states. As a result, when enough incompleteness is combined with noise, while the parabolic shape of performance remains, too much heed tends to be no worse than no heedful communication at all is (see Figures 8(b) and 8(c)).

III. *Novelty.* Until now, we have dealt with agents who have to detect errors, correct erroneous interpretations, and integrate chunks of incomplete information in a world where states of the environment nevertheless correspond to those stored in individual memories. Agents operate under "very trying conditions" (La Porte and Rochlin 1994, p. 221) but still in known worlds. Agents or organizations are often presented, however, with unknown situations (Weick et al. 1999), which they need to interpret. This is the challenge we explore here.

In these simulations, we consider two cases. In the first case (Figure 9(a)), there is a 50% probability that the situation to interpret differs from the ones agents

Figure 9 Novelty



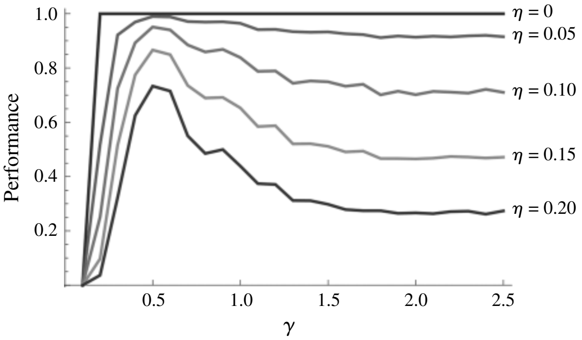
store in memory by three features (a Hamming distance of 3), and a 50% probability that it is one already stored in their memory. In the second case (Figure 9(b)), the probability that the true state of the world is new is 100%. The network of agents interprets the current environment successfully when it correctly discriminates whether the current environment is new or old, and recognizes all features of the environment accurately. We ran 200 iterations for each parameter combination of noise level η and heed γ (5×26). As before, η takes values from 0 to 4 (20% of chance of distortion for each feature) with 5% steps; γ ranges from 0 to 2.5.

Although the results appear to be similar to the ones in Figure 7, there are a few important differences. First, even when noise is absent, the group often fails to interpret the situation when there is no communication because individual memories do not contain the situation to interpret. Once heedful communication is introduced, however, the group becomes good at discriminating between old and new, and it reaches an accuracy close to the performance in the baseline condition of Figure 7(a). Still, peak performance is reached at higher levels of γ than it is in the baseline case: novel environments require more heed than those with only noisy, incomplete information.

Also, although the performance curve again exhibits an inverted-U shape, it reflects different processes than it does in the no novelty condition. In the first trait of the curve, performance is low even when noise is absent because when the situation to interpret is new, agents are "trapped" by their stored knowledge and thus unable to recognize its novelty. This behavior is reminiscent of the famous Mann Gulch disaster. In his rendition of the story, Weick (1993, p. 635) points out that "When the smokejumpers landed at Mann Gulch, they expected to find what they had come to call a 10:00 fire. A 10:00 fire is one that can be surrounded completely and isolated by 10:00 the next morning. The spotters on the aircraft that carried the smokejumpers 'figured the crew would have it under control by 10:00 the next morning' (Maclean, p. 43). People rationalized this image until it was too late. And because they did, less and less of what they saw made sense." In this case, the smokejumpers' prior knowledge, which drove their initial interpretation of the fire, trapped them into an erroneous interpretation that prevented them from seeing that this was not something they were used to. Further, because "when the temperature is approaching a lethal 140 degrees (p. 220), people can neither validate their impressions with a trusted neighbor nor pay close attention to a boss..." Weick (1993, p. 636), they could hardly communicate, which hampered their ability to reinterpret what was going on. Similarly, our results suggest that as heedful communication increases, genuine recognition of novelty emerges and "substitutes" stored memories when needed. Once again, however, consensus on fictive states erodes accurate recognition as heed becomes too strong, and performance degrades. Finally, when all the situations to interpret are novel (Figure 9(b)), peak performance slides further to the right, implying that more heed is needed; the deterioration of performance is significantly attenuated for higher levels of noise. We will discuss this point in the conclusion.

IV. Turnover. Turnover is especially challenging when incoming agents have little in common with either the agents they replace or the other agents in the group. Introducing new agents with different backgrounds can cause a breakdown in the coordination of collective interpretations. Yet, some organizations are resilient to the potentially disruptive effects of turnover (Rochlin et al. 1987). For instance, Michel (2007) studied how newly hired investment bankers performed when they were assigned tasks at once complex and beyond their domain of expertise (for example, the assignment of a merger case to a novice with no banking experience). In the organization Michel (2007) studied, these novices performed consistently well when the organization induced them to communicate intensely with other experts in the bank, and helped them do so through appropriate organizational arrangements. Our simulation considers a similar situation: the case in which one agent out of four is a novice. To emphasize the potential conflict of interpretations, we maximize the difference

Figure 10 Turnover with Heterogeneous Agents



between the newcomers' and incumbents' memories. Figure 10 summarizes the results of 10,000 runs for each parameter combination of noise level η and communication γ. Noise and communication are in the usual range of values.

With low heed, turnover has a strongly negative effect on interpretive performance. Yet, consistent with Michel's findings, heed attenuates such effects. The redundancy that is built into a collective absorbs turnover's disruptive consequences. In line with the prior analyses, Figure 10 also shows that noise negatively affects the amount of absorption achieved. Furthermore, the nonmonotonic effect of heed is apparent, with its usual inverted-U shape, amplified by higher noise levels. Heedful communication thus has beneficial effects vis-à-vis the potential disruptions of turnover, but its benefits disappear for high levels of mutual influence when noise is present.

Discussion and Conclusion

We developed our model with a substantive and a technical goal in mind. We wanted to both foster our understanding of collective interpretation, and offer a tool, an analytical platform that other scholars can use and tailor to analyze a variety of collective interpretive. We now comment on each in turn.

Through our model, we created a simple world in which the phenomenon of interest depends on just a few variables whose behavior and effects we can control precisely. We thus isolated some of collective interpretation's key determinants and studied how they interact to improve or derail interpretation in some relevant configurations. This is what underlies our substantive contribution, which we can understand along three main directives.

First, our analyses tell a different story than is suggested by extant work on collective interpretation. The literature on organizational reliability and collective mind assumes that heed has an overall positive effect, possibly monotonically increasing (more is better). It assumes that collectives that are dysfunctional in their cognitive functions are so largely because people do not pay enough attention to others' states of mind (Walsh 1995, Michel 2007). This statement may be a caricature, but it captures a central tendency of work on collective cognition. By building on a more nuanced characterization of the microfoundation of collective interpretation, our model leads to a partially contrasting perspective. At one level, our analyses clearly support the general result that heed is crucial to interpretive outcomes. We saw that when people pay little attention to others' interpretations, environments that are intrinsically difficult to interpret, either because information is unreliable or unavailable, or because the world is genuinely new, accurate interpretation is hard to achieve. The same can be said when there is turnover. An increase in the degree of heed changes the picture dramatically. When people begin to pay more attention to each other, and to consider others' viewpoints credible, the fog of noise becomes clearer, the darkness of incompleteness lights up, the shape of novel realities coalesces, and the incompetence of novices gets absorbed. At the same time, we also show the dark side of heed. Beyond a certain threshold of heed, the potential for reinforcing erroneous perceptions may be heightened by the excessive weight conferred to others' interpretations. Most prior work has treated the effects of heed over collective outcomes independently. On the one hand, it has emphasized the damages of concurrence seeking; on the other hand, it has noted the virtues of heed. Our closed-form analysis shows how these two aspects are intimately related. Our simulations suggest they interact in a well-patterned way, with their positive and negative effects joining in a parabolic trajectory.

Second, despite the nonmonotonic nature of the relationship between heed and quality of interpretive outcomes, a comparison across the conditions we studied suggests that this relationship is sensitive to the nature of the interpretive task. Two patterns can be identified. One is seen by considering Figures 7–10. In each condition, increased noise causes the performance curves to become steeper: once peak performance is achieved, performance degrades increasingly fast. More noise sharply decreases the collective's tolerance of excessive levels of heed. Too much heed hurts performance, especially when the situation to interpret is ambiguous because information is noisy and unreliable. The second pattern is seen by comparing Figures 7 and 9. This comparison suggests that novelty requires, ceteris paribus, higher levels of heed. That is, for any given level of noise, the collective tends to achieve peak performance for higher levels of heed when it faces a novel situation than it does when it faces a familiar one.[18] For instance, when the noise parameter is set to 0.15 and there is no novelty (Figure 7), the collective achieves peak performance for a γ level of 0.3; when we introduce novelty with probability 1 (Figure 9(b)), the collective achieves peak performance for a γ level of 0.8. We interpret these results in terms of the bright and dark sides of heed. Situations in which information is unreliable offer agents many opportunities to form erroneous perceptions of the environment. An increase in heed exacerbates this pattern, thereby inducing the collective to converge on arbitrary interpretations. Ambiguity ignites the pathological effects of conformity. Conversely, past experiences, knowledge, and histories (the raw materials of associative interpretation) can trap imagination and make it hard for individual interpreters to understand worlds that are distant or genuinely new (March 2006, Gavetti 2012). Collectives can liberate individual minds, and help them understand novel realities as being genuinely novel rather than different manifestations of old realities. This outcome is possible, however, only if individuals weigh others' opinions highly enough: to entertain novel interpretations, individuals need to break the internal consistency of their mental states or memories. To summarize, if we conceive the type of interpretive task along the dimensions of ambiguity and novelty, the general pattern that emerges is one in which the kind of collective that is suited to handle an interpretive task, and how sensitive it is to potentially detrimental variations in heed, depends on the type of task it faces. We are not aware of any work that proposes a contingency approach to collective interpretation. Our analysis does precisely that, thereby introducing a new variable into the design of effective collectives. In this sense, we believe it delineates a fruitful path for future research that empirically explores the key dimensions of such contingency.

Third, although our analyses were inspired by empirical work on high-reliability organizations, we believe they are generalizable to most collective interpretive efforts. Issues of information incompleteness, noisy signals, novelty, and turnover characterize virtually all collective interpretive efforts, from technological innovation to strategic decision making or shop-floor operations. Indeed, the inverted-U-shape-effect of mutual influence, group cohesion, or familiarity has been observed in many empirical contexts. An example that is highly convergent with our analysis is Uzzi and Spiro's (2005) study of Broadway artist networks, which argues that network structures govern behavior by affecting "the level of connectivity and cohesion among actors embedded in the system" (Uzzi and Spiro 2005, p. 449). These are variables whose combined effect would be a good proxy of our parameter γ. Uzzi and Spiro find that connectivity and cohesion are beneficial only up to a certain threshold, beyond which they becomes detrimental. We believe our model can contribute to the theoretical understanding of such nonmonotonic relationships because it examines their underlying cognitive and communicative aspects.

Moving to the technical contribution, the analytical structure we developed is a platform that can be used to address disparate questions for which collective interpretation is central. Indeed, a series of parameters can be readily built into the baseline specification we considered, and Online Appendix 1 shows the model's versatility. Because we think of our model as a platform that other researchers can use, we view our ability to replicate and extend some of the central properties of work on collective mind and organizational reliability as reassuring evidence that the platform we put forth is a reliable basis for studying collective interpretation in other settings. We wish to mention three recent directions in organizational studies that could benefit from our microfounded understanding of collective interpretation, and the analytical platform we offer. First, some recent network studies are paying increasing attention to agency (Burt 2010). Although much of this literature focuses on agents' cognition of the network, the question of how different network structures can be conducive to accurate interpretive efforts fits this literature's broad thrust. Our model can capture different stylized network configurations.[19] Second, recent work on the emergence of organizational forms focuses on audiences that are external to focal organizations as playing a decisive role in whether the emergence occurs (Hannan et al. 2007). In these accounts, as in ours, the interaction among agents who are involved in what are essentially associative interpretive processes underlies how new forms are categorized, and thus whether they eventually gain legitimacy (Carroll and Swaminathan 2000). Again, our model can represent this kind of interplay among audience members. Third, although our model explores different aspects of organizational memory than the research on transactive memory does, it dialogues with this literature by exploring the properties of networks of associative memories. In theories of transactive memory (Wegner 1987) and their applications to organizations (Argote and Moreland 2003, Brandon and Hollingshead 2004, Lewis et al. 2005, Ren et al. 2006), other agents can be "external storages" of knowledge that can be retrieved by locating and retrieving information they maintain. To retrieve given content, one needs to know the address where it is located (a kind of knowledge directory). We model a different process, which is distributed and content-addressable: It is the content that directly triggers memory retrieval from individuals, somehow automatically. Explicit and tacit communication propagates information relevant for activating individual associative memories. The mere act of collective retrieval can actually modify stored memory, as demonstrated by the emergence of novelty. Many forms of collective performance, from classical routines to improvisation in a jazz team (Moorman and Miner 1998), are supported by this kind of content-addressable memory, which allows distributed knowledge to be activated simultaneously. Finally, although our model is not a model of learning, it displays important complementarities with such literature by providing a model of how memory encodes the results of experience. For example, the application of the Hebbian learning rule might be used to model the emergence of patterns of action such as routines and collective habits. It is encouraging that our results tend to converge with recent studies of mutual learning, particularly that of Fang et al. (2010). We see these efforts as complements. Whereas Fang et al. emphasize the effect of small-world type connectivity structure over belief diffusion in processes of mutual learning, our model emphasizes interactions between memories in interpretive tasks, as affected by first order heed rather than connectivity structure in a context in which no learning occurs.

We conclude by pointing out what we view as the main limitation of this article. In our simulation experiments, we created an artificial world, populated by artificial agents, who engage in artificial communication. Although the results we produced are consistent with empirical evidence of collective interpretive systems, the real-life meaning of our parameters remains unspecified. For instance, what is the right level of intensity of communication for a collective that needs to interpret a novel reality? Where is the threshold beyond which a collective falls into a conformity trap? How can it be concretely recognized? These are important questions when it comes to translating our insights into concrete design guidelines. For us, these limitations suggest a clear direction for future research. Together with a more extensive exploration of our model's parameter space, we need empirical work that breathes real life into this space. The journey might not be short, but the payoff can be large.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/orsc.2015.0987.

Acknowledgments

For helpful comments, the authors are grateful to James March, Massimo Marinacci, and Jan Rivkin. The authors also thank Davide Marchiori for computer programming efforts, John Lafkas and Simona Giorgi for excellent research assistance, the Division of Research of Harvard Business School for generous funding, and the European Union's Seventh Framework Programme (FP7/2007-2013) [Grant Agreement No. 318723 (MATHEMACS)]. Errors remain the authors' responsibility.

Endnotes

[1] This is not to say that interpretation is *always* associative. For instance, in a recent illuminating article, Holyoak and Cheng (2011) emphasize the importance of causal learning in certain instances. People do not only interpret the world by drawing associations with past situations, but also by forming understandings of causal relationships based on their observations of events. Although we acknowledge this possibility, our model is designed to capture associative interpretation.

[2] We will use this definition throughout the paper.

[3] Features can be of two different types: specific attributes of the situation, or object features (Tversky 1977), and structural relationships among such attributes, also structural or relational features (Gentner 1983). Revisiting the Charlie Merrill example, *object* features might be characteristics of supermarket customers, attributes of products, or the character of advertising. Typical *structural* feature might be the relationship between the difficulty of product quality assessment and the malleability of consumer tastes. Similarity assessments can involve object features, structural features, or a combination. Although competing theories of associative interpretation have disagreed on how much object features contribute vis-à-vis structural ones in triggering associations, experimental evidence shows that both types of features are important, though individuals tend to focus relatively more on object features (Catrambone 2002). There is 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.

[4] 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).

[5] Neural network models, including the Hopfield model, represent the brain's complex phenomenology only partially. Nevertheless, even their simplest forms seem to capture some basic properties of the brain's functioning beyond what we noted (Hopfield 1982, Hertz et al. 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. 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 (see Kokinov and Petrov 2001) are particularly interesting for their attempt to provide an explicit formal account of structural features in analogy. For our purposes, particularly given our intent to characterize multiagent settings, which implies an

extra layer of analytical complexity, we privilege mathematical tractability and simplicity, and thus focus on the simpler Hopfield model.

[6] Rule (1) represents the standard step update (or activation) function used in most neural network models. In Online Appendix 2 we introduce a continuous sigmoid update function. Online Appendix 2 also compares the behavior of these two activation functions in the case of a collective that faces noisy signals.

[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.

[8] One way to do so (Hertz et al. 1991) involves defining W as the sum of the "outer products" of each stored situation vector s_k with itself. In formal notation: $W = \sum_k s_k \circ s_k$ where $\circ$ is the outer product of two vectors.

[9] Our labeling choice evokes the connections with theories of collective mind in organizations previously discussed (Weick and Roberts 1993). At the same time, our use of heed neglects "higher order" processes emphasized in Weick's and Roberts's use of heedful interrelating, in particular the property of mutual awareness of heed. Indeed, what we model is a sort of "first order heed." We will see in what follows that level of heed is sufficient to generate relevant properties of collective mind, which may suggest significant cautions on generalizations over the virtues of heed that may derive from higher order processes.

[10] A brief digression on the interpretation of the parameter γ is in Online Appendix 3; and a concrete illustration of how collective interpretations are formed in the model in Online Appendix 4.

[11] For this reason, we do not need to provide a proof of Proposition 1, because it is implied by the ordinary proof of existence of fixed points in standard Hopfield nets (see Hertz et al. 1991).

[12] Formal proofs of Propositions 2–4 can be found in Online Appendix 5.

[13] Turnover can also be challenging when group members are heterogeneous and there is a division of cognitive labor (i.e., each agent focuses on some aspects of the environment and pays attention to some select environmental features), and the incoming agents disrupt the pre-established division of labor. We analyzed this situation, but do not include it because of space constraints. Results are available from the authors.

[14] We acknowledge that collective minds have important properties which are related to action feedback that our model does not directly capture (Weick and Roberts 1993); still, we claim that our model can highlight important features of the "constant loop of conversation taking place over several different channels at once" (Rochlin et al. 1987, p. 83) in groups dealing with highly critical environments. Although we agree that feedback from interrelating actions is important, it is still necessary to understand better how different individual minds confer an accurate and coordinated interpretation of the feedback they receive. This is the aspect on which we focus here. Also, similar to theories of collective mind's focus on concepts such as heedful interrelation (Weick and Roberts 1993), we consider the level of mutual attention and influence among agents as the critical parameter determining the properties of collective reliability. As already explained, we summarize such level in a heed parameter γ, which represents how much an agent takes into account the state of mind of the other agents when revising her own beliefs about the current state of the environment.

[15] The number of features is somewhat arbitrary, but allows enough combinatorial possibilities while leaving a manageable computational load.

[16] One way to assess the strength of the γ parameter is to relate it to the probability that for a given γ an actor would reverse her opinion on a specific feature if all other agents would disagree. In the simulation setup of this paper, a γ of 0.5 would imply a reversal of opinion in 11% of the cases, whereas a γ of 1.0 would imply a reversal of opinion in 33% of the cases.

[17] To distract individual memory, retrieval noise must reach a threshold of wrong features sufficient to make the environment signal closer to a wrong prototype than to the right one. The "more than proportional" effect of noise is thus simply due to the amplifying effect of probability multiplication.

[18] In a set of additional analyses, we studied what happens to this relationship for increasingly higher levels of novelty. That is, the current results in Figure 9 reflect a setting in which three features of the situation to interpret are novel with a 50% probability, and a setting in which this probability is 100%. We considered other cases by varying the number of novel features (while preserving these two probability levels). The results we obtained confirmed the pattern suggested above: the higher the novelty, the higher the level of heed that maximizes performance. We do not report such results because of space constraints.

[19] We thank Jim March for suggesting this connection.

References

Amit D, Brunel N, Tsodyks M (1994) Correlations of cortical Hebbian reverberations: Theory versus experiment. *J. Neuroscience* 14(11): 6435–6445.

Anderson JR, Bower GH (1980) *Human Associative Memory: A Brief Edition* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Argote L, Moreland R (2003) Transactive memory in dynamic organizations. Peterson R, Mannix E, eds. *Understanding the Dynamic Organization* (Lawrence Eribaum Associates, Mahwah, NJ), 135–162.

Asch SE (1957) An experimental investigation of group influence. *Sympos. Preventive Soc. Psychiatry*, Walter Reed Army Institute of Research (U.S. Government Printing Office, Washington, DC), 15–17.

Baron RS (2005) So right it's wrong: Groupthink and the ubiquitous nature of polarized group decision-making. *Adv. Experiment. Soc. Psych.* 37:219–253.

Boden MA (1991) *The Creative Mind: Myths and Mechanisms* (Basic Books, New York).

Brandon DP, Hollingshead AB (2004) Transactive memory systems in organizations: Matching tasks, expertise, and people. *Org. Sci.* 15(6):633–644.

Brown TH, Kairiss EW, Keenan CL (1990) Hebbian synapses: Biophysical mechanisms and algorithms. *Annual Rev. Neuroscience* 13:475–511.

Burt RS (2010) *Neighbor Networks: Competitive Advantage Local and Personal* (Oxford University Press, Oxford, UK).

Carroll GR, Swaminathan A (2000) Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry. *Amer. J. Sociol.* 106(3):715–762.

Catrambone R (2002) The effects of surface and structural feature matches on the access of story analogs. *J. Experiment. Psych.: Learning, Memory, and Cognition* 28(2):318–334.

Cohen MD, Levinthal DA, Warglien M (2014) Collective performance: Modeling the interaction of habit-based actions. *Indust. Corporate Change* 23(2):329–360.

Dougherty D (1992) Interpretive barriers to successful product innovation in large firms. *Org. Sci.* 3(2):179–203.

Drazin R, Glynn MA, Kazanjian RK (1999) Multilevel theorizing about creativity in organizations: A sensemaking perspective. *Acad. Management Rev.* 24(2):286–307.

Dutton JE, Dukerich JM (1991) Keeping an eye on the mirror. Image and identity in organizational adaptation. *Acad. Management J.* 34(3):517–554.

Edelman GM (1987) *Neural Darwinism* (Basic Books, New York).

Edelman GM (2006) *Second Nature: Brain Science and Human Knowledge* (Yale University Press, New Haven, CT).

Edmondson AC, Bohmer RM, Pisano GP (2001) Disrupted routines: Team learning and new technology implementation in hospitals. *Admin. Sci. Quart.* 46(4):685–716.

Fang C, Lee J, Schilling MA (2010) Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organ. Sci.* 21(3):625–642.

Feldman MS, Pentland BT (2003) Reconceptualizing organizational routines as a source of flexibility and change. *Admin. Sci. Quart.* 48(1):94–118.

Finkelstein S, Canella B, Hambrick DC (2008) *Strategic Leadership: Theory and Research on Executives, Top Management Teams, and Boards (Strategic Management Series)* (Oxford University Press, USA).

Ford CM, Gioia DA (1995) *Creative Action in Organizations: Ivory Towers Visions and Real World Choices* (Sage Publications, Thousand Oaks, CA).

Fuster JM (1995) *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate* (MIT Press, Cambridge, MA).

Gavetti G (2012) Toward a behavioral theory of strategy. *Organ. Sci.* 23(1):267–285.

Gavetti G, Menon A (2015) Conceptual models of strategic foresight. Unpublished manuscript, Dartmouth College, Hanover, New Hampshire.

Gavetti G, Levinthal DA, Rivkin JW (2005) Strategy making in novel and complex worlds: The power of analogy. *Strategic Management J.* 26(8):691–712.

Gentner D (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Sci.* 7(2):155–170.

Gilboa I, Schmeidler D (2000) Case-based knowledge and induction. *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans* 30(2):85–95.

Gilboa I, Schmeidler D (2001) *A Theory of Case-Based Decisions* (Cambridge University Press, Cambridge, UK).

Glöckner A, Betsch T (2008) Multiple-reason decision making based on automatic processing. *J. Experiment. Psych.: Learning, Memory, and Cognition* 34(5):1055–1075.

Glöckner A, Betsch T, Schindler N (2010) Coherence shifts in probabilistic inference tasks. *J. Behavioral Decision Making* 23(5):439–462.

Goffman E (1974) *Frame Analysis* (Harvard University Press, Cambridge, MA).

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

Hannan MT, Pólos L, Carroll GR (2007) *Logics of Organization Theory* (Princeton University Press, Princeton, NJ).

Hertz J, Krogh A, Palmer RG (1991) *Introduction to the Theory of Neural Computation* (Addison-Wesley Publishing, Redwood City, CA).

Hofstadter DR (2001) Analogy as the core cognition. Gentner D, Holyoak KJ, Kokinov BN, eds. *The Analogical Mind: Perspectives from Cognitive Science* (MIT Press, Cambridge, MA), 499–538.

Holyoak KJ, Cheng PW (2011) Causal learning and inference as a rational process: The new synthesis. *Annual Rev. Psych.* 62: 135–163.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79(8):2554–2558.

Hutchins E (1990) The technology of team navigation. Galegher JR, Kraut RE, Egido C, eds. *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work* (Lawrence Erlbaum Associates, Hillsdale, NJ), 191–220.

Hutchins E (1991) Organizing work by adaptation. *Organ. Sci.* 2(1): 14–39.

Hutchins E (1995) *Cognition in the Wild* (MIT Press, Cambridge, MA).

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

Kandel ER, Schwartz JH, Jessell TM (2000) *Principles of Neural Science*, 4th ed. (McGraw-Hill, New York).

Kokinov BN, Petrov AA (2001) Integrating memory and reasoning in analogy-making: The AMBR model. Gentner D, Holyoak KJ, Kokinov BN eds. *The Analogical Mind: Perspectives from Cognitive Science* (MIT Press, Cambridge, MA), 499–538.

Lakoff G (1987) *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind* (University of Chicago Press, Chicago).

La Porte TR, Rochlin G (1994) A rejoinder to perrow. *J. Contingencies & Crisis Management* 2(4):221–227.

Larkey LB, Markmann AB (2005) Processes of similarity judgment. *Cognitive Sci.* 29(6):1061–1076.

Lewis K, Lange D, Gillis L (2005) Transactive memory systems, learning, and learning transfer. *Organ. Sci.* 16(6):581–598.

March JG (2006) Rationality, foolishness, and adaptive intelligence. *Strategic Management J.* 27(3):201–214.

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

McRae K, de Sa VR, Seidenberg MS (1997) On the nature and scope of featural representations of word meaning. *J. Experiment. Psych.* 126(2):99–130.

Meindl JR, Stubbart C, Porac JF (1996) *Cognition Within and Between Organizations* (Sage, Thousand Oaks, CA).

Michel AA (2007) A distributed cognition perspective on newcomers' change processes: The management of cognitive uncertainty in two investment banks. *Admin. Sci. Quart.* 52(4):507–557.

Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335(6193): 817–820.

Moorman C, Miner AS (1998) Organizational improvisation and organizational memory. *Acad. Management Rev.* 23(4):698–723.

Mullainathan S, Schwartzstein J, Shleifer A (2008) Coarse thinking and persuasion. *Quart. J. Econom.* 123(2):577–619.

Narduzzo A, Rocco E, Warglien M (2000) Talking about routines in the field: The emergence of organizational capabilities in a new cellular phone network company. Dosi G, Nelson RR, Winter SG, eds. *The Nature and Dynamics of Organizational Capabilities* (Oxford University Press, New York), 27–50.

Neustadt RE, May ER (1986) *Thinking in Time: The Uses of History for Decision-Makers* (Free Press, New York).

Obstfeld D (2012) Creative projects: A less routine approach toward getting new things done. *Organ. Sci.* 23(6):1571–1592.

Perkins EJ (1999) *Wall Street to Main Street: Charles Merrill and Middle-Class Investors* (Cambridge University Press, Cambridge, UK).

Pinker S, Prince A (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1):73–193.

Porac JF, Thomas H, Baden-Fuller C (1989) Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers. *J. Management Stud.* 26(4):397–416.

Poucet B, Save E (2005) Attractors in memory. *Science* 308(5723): 799–800.

Rawls J (1971) *A Theory of Justice* (Belknap Press of Harvard University Press, Cambridge, MA).

Ren Y, Carley KM, Argote L (2006) The contingent effects of transactive memory: When is it more beneficial to know what others know? *Management Sci.* 52(5):671–682.

Resnick LB, Säljö R, Pontecorvo C, Burge B (1997) *Discourse, Tools, and Reasoning: Essays on Situated Cognition* (Springer, Berlin).

Rochlin GI, La Porte TR, Roberts KH (1987) The self-designing high-reliability organisation: Aircraft carrier operations at sea. *Naval War College Rev.* 40(4):79–90.

Rosch E (1978) Prototype classification and logical classification: The two systems. Scholnick EK, ed. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?* (Lawrence Erlbaum Associates, Hillsdale, NJ), 73–86.

Rumelhart DE (1992) Toward a microstructural account of human reasoning. Davis S, ed. *Connectionism: Theory and Practice* (Oxford University Press, New York), 69–83.

Rumelhart DE, Ortony A (1977) The representation of knowledge in memory. Anderson RC, Spiro RJ, Montague WE, eds. *Schooling and the Acquisition of Knowledge* (Lawrence Erlbaum Associates, Hillsdale, NJ), 99–135.

Sandelands LE, Stablein RE (1987) The concept of organization mind. Bacharach S, DiTomaso N, eds. *Research in the Sociology of Organizations* (Jai Press, Greenwich, CT), 135–161.

Smith WK, Binns A, Tushman ML (2010) Complex business models: Managing strategic paradoxes simultaneously. *Long Range Plannins* 43(2–3):448–461.

Smolensky P, Legendre G (2006) *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, Vol. 1: Cognitive Architecture (MIT Press, Cambridge, MA).

Thagard P (2014) Cognitive science. Zalta EN, ed. *The Stanford Encyclopedia of Philosophy* (Fall 2014 Ed.), http://plato.stanford.edu/archives/fall2014/entries/cognitive-science/.

Tversky A (1977) Features of similarity. *Psych. Rev.* 84(4): 327–352.

Uzzi B, Spiro J (2005) Collaboration and creativity: The small world problem. *Amer. J. Sociol.* 111(2):447–504.

Volkema RH, Farquhar K, Bergmann TJ (1996) Third-party sensemaking in interpersonal conflicts at work: A theoretical framework. *Human Relations* 49(11):1437–1454.

Walsh JP (1995) Managerial and organizational cognition: Notes from a trip down memory lane. *Organ. Sci.* 6(3):280–321.

Wegner DM (1987) Transactive memory: A contemporary analysis of the group mind. Mullen B, Goethals GR, eds. *Theories of Group Behavior* (Springer-Verlag, New York), 185–205.

Wegner DM, Erber R, Raymond P (1991) Transactive memory in close relationships. *J. Personality Soc. Psych.* 61(6):923–929.

Wegner DM, Giuliano T, Hertel PT (1985) Cognitive interdependence in close relationships. Ickes WJ, ed. *Compatible and Incompatible Relationships* (Springer-Verlag, New York), 253–276.

Weick KE (1990) The vulnerable system: An analysis of the Tenerife air disaster. *J. Management* 16(3):571–593.

Weick KE (1993) The collapse of sensemaking in organizations: The mann Gulch disaster. *Admin. Sci. Quart.* 38(4):628–652.

Weick KE, Roberts KH (1993) Collective mind in organizations: Heedful interrelating on flight decks. *Admin. Sci. Quart.* 38(3):357–381.

Weick KE, Sutcliffe KM, Obstfeld D (1999) Organizing for high reliability: Processes of collective mindfulness. *Research in Organizational Behavior* (JAI Press, Greenwich, CT), 81–124.

Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. *Science* 308(5723):873–876.

Giovanni Gavetti is an associate professor of Business Administration at the Tuck School of Business at Dartmouth. He received his B.A. in Economics from Bocconi University in Milan, and his M.A. and Ph.D. in Management from the Wharton School, University of Pennsylvania. His research focuses on the cognitive foundations of strategy.

Massimo Warglien is a professor at the Department of Management of the Ca' Foscari University of Venice, where he is also the director of the Center for Experimental Research in Management and Economics. His research interests include strategic thinking, organizational adaptation, language games, and the perception of time.