Counter-Terrorism as Neighborhood Watch:
A Socio/Computational Approach for Getting Patterns From Dots

Max Boisot

Universitat Oberta de Catalunya, Barcelona, Spain
The Sol Snider Center, The Wharton School, University of Pennsylvania.
tel: (34) 93 894 8915 fax: (34) 93 894 9223 boisot@attglobal.net

&

Bill McKelvey

The UCLAAnderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095-1481
tel 310-825-7796 fax 310-206-2002 mckelvey@anderson.ucla.edu

Draft 20 December 18, 2004

© All rights reserved.
Not to be quoted, paraphrased, copied, or distributed in any fashion.

ABSTRACT

Improved security, it has been argued, consists in getting more dots and in the joining them up in a timely manner. In this paper we argue that such an approach rapidly leads to computationally intractable intelligence tasks and does not actually deliver what is needed. We take *timely pattern processing* to be the main intelligence task—computationally speaking, an even more daunting prospect. After discussing the nature of the problem in the light of Ashby's Law of Requisite Variety, we introduce the concept of Global Neighborhood Watch as a socially distributed pattern processing strategy. We show how the application of filters that privilege certain contexts, vantage points, and time frames, in conjunction with a distributed intelligence process drawing on modern "citizen" technologies can rapidly home in on relevant security threats. We illustrate the process by means of an imaginary case and discuss the policy implications that it raises. A conclusion follows.

What role might a citizen play in combating terrorism? Consider the following scenario:

> The year is 2010. You are traveling by subway from Dulles International Airport to Union Station in Washington. The time is 9:20am; the train is moderately crowded. You are hanging onto the overhead handgrip when your cell phone emits a faint beeping sound. It does so because a chemical sensor in your computer bag has detected a possible presence of a PETN-based material in the subway car (a chemical used in the manufacture of sophisticated form of C4, the kind of explosive Richard Reid hid in his shoes when he boarded American Airlines Flt. 63 out of Paris in December 2001). You start looking around. You spot a beat-up looking rucksack under a seat at the far end of the car. You wonder whether you should act. It is a difficult decision. Your beeper has occasionally sounded in the past, as you have traveled here and there. By questioning passengers, you could end up making a fool of yourself. You might also be revealing that you are a member of the Washington chapter of *Global Neighborhood Watch* (GNW), an international citizen-inspired antiterrorist network. Members of the US branch of the network are required to respond to signals and report back to the National Counter-Terrorism Center (NCTC) that was created in 2005. For the time being you decide to do nothing, knowing that your beeper has already fired off a signal to the NCTC's computer. At the next stop, two women board at the far end of the car. One of them is also a member of GNW. Now her beeper also goes off—with a more intense beeper code. The NCTC's computer "sees" that the two signals come from the same cell district. It signals back to the beeper units with a code asking you each to independently "text-message" back your location and what you see; or, you and the other GNW member immediately receive a phone call from an agent at the NCTC. You each describe the situation—independently of each other.[1]

As a member of GNW, you are an intelligent communication node embedded in a globally distributed network of sophisticated sensors. Your first responsibility when your beeper goes off is to register the *possibility* of a threat. Once this has been established, the next step is to assess its *plausibility*. For this you need further corroborative evidence. In our subway example, this was provided by the second GNW member's beeper also firing—at a greater intensity. In effect, she provided NCTC with a Bayesian update on what was initially a fairly low a priori probability estimate. If your beeper had gone off when you were alone in the middle of a field, for example, the degree of corroboration required to establish plausibility would have been higher than for a subway.

Its faintly futuristic quality apart, how realistic is the above scenario? In an invited article in *The Economist* commenting on the findings of the US and UK commissions of inquiry, Efraim Halevy (2004: 21), former head of Mossad, Israel's intelligence service, wrote the following:

> There is an inherent understanding…that the shortage of information on the threats—from Islamic terrorism and from Iraq's weapons of mass destruction—was at the root of the intelligence breakdown on these two fronts. It seems only logical that the more you know, the safer you are and the greater the chance that you will get things right.
>
> Yet Israel's most costly and fateful failure was its mistaken estimate of Egyptian and Syrian intentions, on the eve of the Yom Kippur war in 1973, when the two armies unexpectedly attacked Israel in a bid to regain the territories lost in the 1967 war. At the time, Israel had it all: superior intelligence coverage, excellent human resources with good access, high-level and discreet dialogue with more than one Arab or Muslim leader, and an intelligence-evaluation arm that had provided an early warming several months before the war, thus preventing it from breaking out at that time. But despite all of the above, we got it all wrong. The abundance of information led us to intelligence "hubris": we trusted our superior analytical prowess rather than ominous indicators on the ground."

Clearly, the sufficiency of information has not been the only issue. The ability to rapidly link different items of information together has also been recognized as important. To this end, the *Final Report* of the September 11th Commission (2004), for example, proposed the creation of the National Counter-Terrorism Center that would concentrate all the analysts and spies working on counter-terrorism in one place. The September 11th attacks were made all the easier, argues the Commission, because not enough information was shared among the agencies. The proposed new center "…will help to connect the dots" (*The Economist* 2004: 30). Yet the CIA emphasized that it had warned senior policy-makers of the terrorist threat well before it happened: "Our fundamental flaw was not withholding information. The flaw was in not recognizing the significance of the information that we did have and acting on it promptly" one senior official said (*The International Herald Tribune* 2004:). In both the US and the UK, intelligence assessment is acknowledged to have been as much of a problem as intelligence gathering (*Financial Times* 2004).

In this paper, we explore the nature of the data-processing challenges and opportunities that the new NCTC will confront in the modern age. Empowered by modern information and communications

technology, asymmetric threats operate a level of complexity that intelligence organizations have not had to deal with hitherto. Such complexity can either be *reduced* (a cognitive strategy), or *absorbed* (an organizational one) (Boisot 1998, Boisot & Child 1999). We shall deal with complexity reduction first. In the next Section we first outline the much discussed but little understood distinction between data information and knowledge. In Section 2, we offer a new interpretation of Ashby's law of requisite variety and in Section 3 we frame the problem of complexity reduction as one of cognitive filtering—moving from possible to probable hypotheses while satisfying plausibility constraints. In Section 4 we illustrate our approach by further elaborating the scenario that we presented above. We discuss the organizational implications of our analysis in Section 5 and then offer a brief conclusion.

1 THE CHALLENGE OF COMPLEXITY REDUCTION

1.1 DATA, INFORMATION, AND KNOWLEDGE

Any living creature survives by first scanning its environment for data on the threats and opportunities that confront it. It must then extract the relevant information from such data and interpret it according to its prior experience before acting on it. Where it scans from depends on what data it is sensitive to. This defines its *Umwelt* (von Uexküll, 1921)—that is, those parts of the real world (ontology) it has direct sensory access to. In the case of humans, the range of data that we are sensitive to has been considerably expanded by technologies that allow us to capture the very small via microscopes or the very large via telescopes. These technologies often create problems of data overload that require us—now aided by computer imaging—to separate out what is informative data from what is noise. Effective scanning thus calls for selectivity with respect to what will count as data, what information will be extracted from it, and what information will be acted upon. Such selectivity is provided by a set of filters that are tuned by past experiences to establish what will validly count as data and what information will be extracted (Figure 1).

<<< Insert Figure 1 about here >>>

Figure 1 highlights the relationship between *data*, *information* and *knowledge* (Boisot & Canals 2004). Any agent operates a set of tunable filters, both perceptual and conceptual that allow past knowledge, values, and preferences to shape both what will register as data as well as what information will be extracted from such data. Information then modifies an agent's knowledge, viewed as a set of probability distributions or conditional expectations (Arrow 1984). We can think of these filters as expressing a set of hypotheses as to what to look for in environmental signals. As indicated in Figure 1, agents have two kinds of action options: (1) *Epistemic (cognitive) action* tunes perceptual and conceptual filters, based on hypotheses—mental constructions or "ideas that die in our stead" as Popper (1972) put it—that agents hold about the world; (2) *Pragmatic (behavioral) action* acts directly upon the world and, in so doing, indirectly modifies incoming signals prior to their being filtered. An agent, for our purposes, can be any living system—amoeba, human, organizational, etc—that exhibits both coherence and intentionality (Thagard 2000).

1.2 DOTS, LINKS, AND PATTERNS

The tunable filters that shape our hypotheses can sometimes function too well. When confronted with a radically novel situation, relying solely on past experiences to tune the filters can end up unduly narrowing the compass of one's scanning efforts. The dilemma is clear: Admit too much data and you confront overload. Admit too little and you miss either the novel threat or the novel opportunity. This dilemma is visible both in the tragedy of 9/11 and in the 2003 Iraq war. The 9/11 threat was missed because the intelligence services' tunable filters were attuned to the wrong data—i.e., their scope was too narrow. The 2003 Iraq war, by contrast, was a response to threats that turned out not to be there. In each case the challenge was to come up with an appropriate epistemic strategy in a timely manner, one undistorted by political interests.

What, in effect, has been the response to asymmetric threats of the Al-Qa'ida type? Writing in *The Economist* (2003: 30), six of America's most experienced intelligence practitioners argued that although there had been "…an inability to connect the dots" what is really needed is more useful dots to connect, more fine-grained and better quality data, and more monitoring based on the data. We believe, however,

that *simply pleading for more and better dots is to mistake the nature of the problem*. An arithmetic increase in the number of "dots" to play with—high quality or otherwise—leads to a geometric increase in the possible connections that one can establish between them. It also leads to an exponential increase in the number of patterns that can be generated from connected dots. Consider Table 1 below:

Number of dots: N	Number of possible links: $L = N(N–1)/2$	Number of possible patterns: $P = 2^L$
$N = 4$	$L = 6$	$P = 64$
$N = 10$	$L = 45$	$P = 3.5$ trillion
$N = 12$	$L = 66$	$P = 4{,}700$ quadrillion

Table 1. Relation of dots to patterns

Four dots lead to six possible links; they also generate 64 possible patterns. Add 6 dots and one gets 10 dots generating 45 possible links and approximately 3.5 trillion possible patterns. Add just 2 more dots and you are dealing with 66 possible links and approximately 4,700 quadrillion possible patterns! The implication of this progression is that whatever we feel about the need for more high quality dots to connect—and we are not denying that the need is a real one—if we are not to drown in a sea of unprocessable data, we need to home in on meaningful patterns while simultaneously filtering out the much larger number of those that are meaningless.

The cognitive approach to variety reduction brings us back to the distinction between data, information, and knowledge. Simplifying somewhat, dots are environmental signals that register with an agent as *data*—a beeper going off, a rucksack under a seat, a moderately crowded subway car, etc. Links between dots constitute *information* that an agent extracts from the data, and the patterns that can be derived from such information when properly filtered can become *knowledge* that the agent can act upon. Patterns, however, only acquire the status of knowledge for an agent when they have graduated from being merely *possible* to being sufficiently *probable* to justify action. Note that, in contrast to the Platonic definition of knowledge as "justified true belief," such a pragmatic definition of knowledge does not require it to be certain—i.e., true—but merely probable enough to justify action.

Data, however, are not patterns and nor, indeed, are dots. The problem is well known in science. The Duhem-Quine thesis holds that we never test hypotheses in isolation but in patterned clusters or chunks (Duhem 1914, Quine 1969). But as we have just seen, from a few dots and links a huge number of patterns become candidates for testing. The patterns that survive the testing—i.e., that are now judged to be either "plausible" or "probable"—will subsequently orient us in an iterative fashion towards attending to particular dots and particular links now deemed to be relevant (Hanson 1958). Yet, given the unimaginably large number of patterns that we have to start with, how does the process ever get going? *How do we winnow out unviable patterns and process the rest in a time frame that allows for efficacious action and adaptation?*

The issue, a familiar one in mathematics and computer science, is one of computational tractability. As we have just seen, while the number of data points increases arithmetically, the number of possible patterns that can be generated from these increases exponentially. Thus, while in many circumstances we face an exponential growth of problem size, at best our data processing and transmission capacities can cope with a polynomial growth of problem size. It turns out, however, that we humans are fairly good at processing patterns, deploying those that we have accumulated almost unconsciously through prior experience to make sense of the world (see Figure 1). We tend, however, to formulate hypotheses in the neighborhood of our past experiences and these, then, act as filters for making sense of the new experiences that we encounter. In relatively stable environments this makes efficient use of scarce data processing resources. But we often find it difficult to move out of the old neighborhood when the novelty of the circumstances requires it, or when, because of technological progress, data processing resources suddenly increase. Such moves carry a re-location cost. This is the well-known problem of paradigm change first identified and discussed by Kuhn (1962). Here, the challenge is to improve one's pattern recognition skills rather than one's data collection skills. If one's existing repertoire of patterns or paradigms is inadequate for interpreting incoming data,

then it has the effect of converting information-bearing data into noise. Collecting more data under such circumstances ends up exacerbating the problem since it merely increases the level of noise.

But what does pattern recognition involve? It requires matching the patterns suggested by incoming data to stored templates. Do we, therefore, need to hold in memory quadrillions of pattern templates to cope with the whole range of possible patterns generated by the data we collect? No! We adopt the strategy of our immune system (Kauffman 1993: Ch. 4) and exploit the combinatorial power of a limited number of pattern elements already in our repertoire. Providing we can mix and match these flexibly and rapidly to incoming data, we greatly reduce the number of patterns that we need to store.

Even so, such pattern reducing strategies may still not be enough. Our own memory, its combinatorial powers notwithstanding, may still end up filtering out as meaningless, patterns that are potentially significant. In effect, we run the danger of reducing complexity too quickly. One way of improving our pattern recognition skills is to harness those of other agents so as to create a distributed processing capacity. This is what the institutions of science are designed to do. In spite of its competitiveness, science is essentially a collaborative enterprise in which all players share a concern to extract meaningful information from data in order to come up with novel patterns—a form of socially distributed information processing.

When dealing with security threats, however, and in contrast to the normal case in science, key players have an adversarial rather than a collaborative relationship with each other, one that is best handled from a game-theoretic perspective (Binmore 1994). While some players wish to extract information from data, their adversaries want to hide information in data—this is the logic of immunology operating at the human scale. Here, we are in effect dealing with a form of encryption, a game of hide-and-seek played with information. If *de*cryption is about extracting useful information from data, *en*cryption is about hiding useful information in data. In the absence of effective decryption strategies—filtering, by any other name—collecting more data under such circumstances merely exacerbates the problem since it ends up making it easier to hide information rather than more difficult. Hiding information in noise increases the variety one has to respond to, a problem aptly captured by the cybernetician, Ross Ashby.

2 ASHBY'S LEGACY

2.1 ASHBY'S LAW

Ashby's Law of Requisite Variety states that *only variety can destroy variety* (Ashby 1956: 207). More specifically, his law holds that for a biological or social entity to be efficaciously adaptive, the variety of its internal order must match the variety of the environmental constraints that it confronts. In defining variety, Ashby (1956: 124–25) pointed to the following series: "*c*, *b*, *c*, *a*, *c*, *c*, *a*, *b*, *c*, *b*, *b*, *a*." He observed that *a*, *b*, and *c* repeat, meaning that there are only three "*distinct* elements"—three kinds of variety or three degrees of freedom. In the language of patterns, however, this is variety at the level of "dots." Suppose, instead, we define variety in terms of the number of patterns instead of the number of dots. Then, using the formulae from Table 1, we see that 12 dots allow 7,400 quadrillion possible patterns—a large amount of variety to be destroyed! Even supposing 99% of these are not worth paying attention to, trillions are left, and one still doesn't know, up front, which ones are trivial and which are not.

Since variety is but the phenomenological manifestation of complexity at work, we extend Ashby's treatment of variety to complexity and argue that *only complexity can destroy complexity* (McKelvey & Boisot 2003). How might we apply our reformulation of Ashby's law? Faced with an external stimulus, a living creature can essentially marshal two types of reactions: (1) a purely behavioral response of the stimulus-response variety, one that is unreflectively based on instinct; or (2) an intelligent and reflective response based on an interpretation of the stimulus, one that interposes itself between the stimulus and the behavior that it gives rise to. The first type of response is hard-wired and requires an adaptive response to be already available in the repertoire to match whatever stimuli the creature encounters. The second type of response exploits the creature's "reprogrammable software"—i.e., its intelligence—to reduce the amount of hard wiring required. In effect, as already indicated in Figure 1 above, the second type of response simultaneously modifies what counts as meaningful external variety and what counts as a meaningful internal response. With respect to external stimuli, the creature's strategy is to filter out noise, that is, data that are not information bearing. Internally, its strategy is to exploit the combinatorial powers of its

response repertoire in order to come up with an adaptive response without exhausting its energy resources. In the case of human beings confronting ever more complex environments, both types of response confront cognitive and behavioral biases resulting from an improper use of prior experiences and routines, from excessive stress (Staw, Sandelands & Sutton 2003), and from groupthink (Janis 1972). The two kinds of biases mutually reinforce each other.

We illustrate the above points in an *Ashby Space*, shown in Figure 2. On the vertical scale we measure the complexity of the stimuli impinging on a living system from outside. On the horizontal scale we measure the internal complexity of the living system's response to such external stimuli. The 45º line and below indicates the region in which the complexity of the response is at least as great as that of the stimulus. Complexity on the diagonal or below it is thus *requisite*. Above the 45º line, however, at, say, point A, the complexity of the stimulus exceeds that of the response. In such circumstances, two strategies are possible: (1) either increase response complexity by moving horizontally to the right from point A to point B in the diagram; or (2) reduce stimulus complexity by moving vertically down from point A to point C in the diagram. As indicated by line AD in the diagram, the two strategies are typically combined. The first strategy, AB, we take to be primarily *behavioral* involving little thought but much expenditure of scarce energetic resources; the second strategy, AC, we take to be primarily *cognitive* and economic in its use of scarce energetic resources. A "mindless"—i.e., purely behavioral—response to external stimuli can lead to an excessive expenditure of energy and to the ultimate disintegration of a living creature. Since we are interested in the application of intelligence to external threats and opportunities, we shall focus on the cognitive response to complexity.

<<< Insert Figure 2 about here >>>

Stimuli are noisy data that register on the sensory apparatus of a living system. The challenge is that of extracting information from the data and discarding non-information bearing data—that is random noise—in a timely fashion. Moving down the vertical scale of the Ashby Space reduces "noisy" complexity by focusing on the regularities that reside in the stimuli. More importantly, moving down the space economizes on unnecessary behaviors—and hence on scarce energetic resources—by reducing the distance that one has to move to the right before encountering the 45º line—say at point D in the diagram.

The problem is that we live in a dynamic world. As phenomena evolve, so do the data by which they manifest themselves. In the early phases of emergent order creation, for example—processes by which new patterns evolve—the data are ambiguous and point to myriad alternative possibilities. Over time, however, as certain states of nature are "selected" and begin to stabilize, they show their hand in the data. The perceived ambiguity of phenomena then gradually reduces. Out of the myriad possibilities, certain types of pattern now *resonate* with those that a living creature can construct from its limited repertoire. These, then, become more *plausible* than competing alternatives. With the further passage of time and the arrival of further confirmatory stimuli, a subset of these plausible patterns will now strike the creature as distinctly *probable*. At the end of the process, one state is now seen or experienced as being sufficiently likely to justify one of a limited number of responses—such as fight or flight. When the state is experienced as certain, it eliminates all other contenders and only one response is needed. When the state is merely experienced as highly likely, however, it may give rise to a range of possible responses designed to cover several contingencies. These then look more like the taking out of an insurance policy or an option.

In this way, an intelligent creature typically passes from possible to plausible and then on to probable worlds—under certain circumstances an alternative move from possible to probable worlds and then onto plausible worlds also happens (Boisot & MacMillan 2004). Much depends on the repeatability and intelligibility of events. Either way, the actual world, the one in which certainty obtains, only shows up at the end of the process, and more often than not, once the creature's response—adapted or not—is already in place. In this paper, we focus on the *possible-plausible-probable-actual* trajectory. The journey from possible worlds to the actual world registers as phase transitions as one moves down the vertical axis of the Ashby Space (see Figure 3). An intelligent response to the law of requisite complexity involves moving as far down the vertical scale as one can, sufficiently ahead of events so as to minimize the complexity of an adapted response—and hence minimize the energy expenditures necessary to deal with them. Yet one can only move down as far as one's *umwelt*—the ontology one has access to—allows. In the case of human

beings, as we have seen, the ontology has been extended by science and technology to cover worlds—both large and small—that are inaccessible to the human senses on their own.

<<< Insert Figure 3 about here >>>

One can move down the Ashby space too slowly or too fast to be appropriately adaptive. The requirement for certainty slows the downward movement. But under threat, the danger is that one will move down too quickly. The intelligence failures of 9/11 and of the Iraq war are, respectively, examples of just such slow or over-hasty moves down the vertical scale of Figure 3. In the case of 9/11, given the tensions created by the terrorist threat, complexity was reduced too slowly. In the case of the Iraq war and the search for weapons of mass destruction, it was reduced too fast. Any move from possible to plausible worlds and from these into probable worlds should be *truth improving*, gradually eliminating unviable beliefs and generating beliefs that one could act upon—i.e., knowledge. The truth to be improved is made up of two components:

- ✓ ***Coherence:*** One tentatively accepts a hypothesis as true because it offers intelligible understanding of the phenomenon under investigation—i.e., it resonates with personal experience (BonJour 1985) while meeting the test of logical consistency.
- ✓ ***Correspondence****:* A hypothesis corresponds to phenomena that can be observed in the external world as, for example, established by standard, sequential (statistics-based) research over time (Tarski 1956).

Coherence and correspondence approaches to truth are always intertwined, although at different moments one will predominate (Boisot & MacMillan 2004). The move from possible to plausible worlds is based on *coherence*—based on one's past experiences, do the parts of the world one is aware of fit together and make sense? Does it paint a logically consistent picture? A coherent one? The move from plausible to probable worlds, by contrast is based on *correspondence*—does the pattern one sees correspond to what is actually out there? In sum, if possible worlds refer to worlds that we can conceive of, plausible worlds constitute a subset of possible worlds, those that are coherent enough to make sense to us. Probable worlds constitute a second subset of possible worlds. Where the two subsets intersect we obtain a world for which we can legitimately formulate future expectations. As we have already seen, we can get from possible worlds to the intersection of the two subsets either via plausible worlds or via probable ones (Boisot and McMillan, 2004). Probability, however, implies repeatability. And since the events that we are discussing in this paper tend to be singular in nature, we shall choose to reach the intersection via plausible worlds.

But how does one generate plausible hypotheses? And, having done so, how does one then select from these hypotheses that are probable? The challenge: *How to extract meaningful information from noisy data in a timely fashion while maintaining a truth-improving stance?* Figure 1 tells us that a filtering process is involved, one that is shaped by prior expectations and preferences. These activate tunable filters that drive the search for weak signals and their interpretation. As the Figure indicates, through pragmatic action, there is also an enactment of possible outcomes that can shape the world we are called to respond to. In the next Section we explore the move from possible to probable worlds via plausible ones.

3 GOING FROM POSSIBLE TO THE PROBABLE VIA PLAUSIBLE WORLDS

3.1 USING CONTEXT TO COARSE-GRAIN FINE-SCALE STRUCTURE

The concept of the Global Neighborhood Watch (GNW), that we presented in the Introduction, builds on the insight that as individuals in a social network interact over time, they influence each other's behavior, if only in some limited way—they exchange things and ideas; have agreements and disagreements; discover common interests, likes, dislikes, shared attitudes, values and prejudices, share past experiences, etc. Like elementary particles in quantum theory, therefore, individuals carry in memory a *history* of their interactions with all other individuals they have come into contact with over time—that is, pairs or larger groupings of individuals exhibit *correlated social histories* (McKelvey 2003, 2004b). We will use the term "*transaction*," as opposed to "interaction," to reflect our interest in meaningful rather than neutral encounters between individuals. As transactions multiply, a person's behavior can be inferred with some probability from the behavior of those s/he has transacted with—i.e., through their correlated social histories (one for each person transacted with). Thus, instead of describing a person's behavior directly—this may, after all, be inaccessible to direct observation—we could do what quantum theorists do and derive it from his/her correlated histories with all other people in his/her network.

In quantum theory, *entanglement* occurs when the *correlated histories* of pairs of electrons are greater than zero. If individual histories are thus correlated, they are said to *interfere* with each other. Correlated histories across a random collection of heterogeneous individuals will have their predictive effects canceled out due to the randomness of the forces at play. Gell-Mann (1994: Ch. 11) refers to a world densely populated with interference-prone histories as having "*fine-grained*" structures. The quantum world is just such a fine-grained structure. By contrast, Gell-Mann labels the kinds of patterns that we see in the material world as *coarse-grained* structures. The question then arises: *How do coarse-grained patterns emerge from fine-grained—entangled—structures?* More broadly: how does pattern emerge from a background of everything-more-or-less-correlated-with-everything-else that cancels out?

This is equivalent to asking: *How can we obtain meaningful patterns from connected dots?* According to Gell-Mann, "A coarse-grained history may be regarded as a class of alternative fine-grained histories, all of which agree on a particular account of what is followed, but vary over all possible behaviors of what is not followed, what is summed over" (p. 144). Researchers exploit this phenomenon every time they assume that the various interrelated effects not specifically hypothesized, or controlled for, are randomized, neutralize each other, and summed over. Fine-grained structures incorporate all the error terms.

Gell-Mann's view is that *contextual effects* lead some correlated histories in the fine-grained structure to get selected as the basis of *probabilistic patterns* while the remaining histories are washed out—their effects remaining randomized (see also Omnès 1999). For example, faced with some urgent, external work-related problem, a project team's decision process focuses on the contextually imposed problem and *not* on all the many other individual concerns of its members. This is context-driven coarse-graining emerging from the fine-scale structure of all-possible concerns and correlated social histories of the various members. Gell-Mann's "Modern Interpretation" of quantum theory, starts with the most basic probability formation process in physics, how quanta get transformed into visible matter. Omnès emphasizes the role of external context. We use their redefinitions of quantum theory to explain how *probable* patterns emerge from *possible* ones. For us, these are the first and most rudimentary pattern-forming processes. From our "dot" analysis, we know that an almost infinite number of possible patterns can be derived from a modest number of dots. We use thinking from quantum physics to guide our thinking about how patterns first emerge and to show how moving from *possible* to *probable* patterns via *plausible* ones can reduce pattern proliferation.

3.2 THREE PATTERN FILTERS

The first step is to reduce the trillions of possible patterns "out there" at as low a cost as possible. While 12 dots produce 66 links, these pale beside the 2^{66} possible patterns that could emerge from these links. As noted above, it is not "filling in the (missing) dots" that is the problem. Instead, given even a relatively few dots, it is the unmanageably large number of patterns emerging from the "fine-grained structure" of *connected* dots that initially do nothing more than obscure the "coarse-grained" possibility of some pending catastrophe.

In our scenario a dot is just a data point—a spatio-temporally situated state of nature existing at some given level of complexity. It could range from "*A person just entering my subway car may have set off my beeper*" to "*A white middle-class woman wearing green sunglasses and wearing a black leather coat bulging at mid-torso has entered my car carrying a strange-looking orange bag; my beeper was triggered two seconds after she entered*." The second "dot" is clearly more information-bearing than the first—it has more *attributes*. The basic pattern-generating matrix thus, consists of $\boldsymbol{d}$ dots × $\boldsymbol{h}$ attributes. So, the number of possible patterns possibilities may actually turn out to be far higher than the 2^{66} derivable from 12 dots. Each "dot" turns out to be an $\boldsymbol{h}$ attribute vector that can connect at one or more points with the vectors associated with other dots in a connectivity matrix. Clearly, the more points at which two vectors connect on, the stronger the correlated histories of their respective dots, and, by implication, the stronger the connection between them. For example, was the "white middle class women wearing a black leather coat" at school with the Malay-looking young man who entered car number 7 two stops before? He was also carrying a strange-looking orange bag and also triggered the firing of a beeper. None of this, however, necessarily constitutes a threat. No one knows. Thus, all of the dot-vectors *interfere* with each other to produce *entanglement*, while concealing from outsiders the connections of possible interest.

Any person may link with another person with respect to each attribute—sunglasses to sunglasses, common schools, orange bags, (perhaps even similar contents?), etc. At the maximum, we could see 2^{66} pattern-element connections *times* the $\boldsymbol{h}$ attributes comprising each vector = $2\boldsymbol{h}^{66}$. For some large number of agents, there would be many dissimilar or non-complementary attributes, so in reality a given dot-vector, $\boldsymbol{x}$, is a long bit-string containing some 1s and mostly 0s with occasional 1s at the same level as, say, vector $\boldsymbol{y}$—where some association *could* be established. Possibly no linking occurs at all, however, so that the number of patterns remains zero. The basic array of $2\boldsymbol{h}^{N(N-1)/2}$ *possible* patterns could be vast indeed but most pattern-elements (or frequencies) will cancel out and disappear into the fine-scale structure.

We now propose a three-step filtering approach to pattern processing, one for each of the possible, plausible, and probable worlds. The operation needs to be computationally tractable and performed at a faster rate than, say, the rate at which a given threat or opportunity builds up. Our proposed hi-tech "socio-computational" approach follows in Section 4.

3.2.1 Step 1: Identify *Contextual Tensions* and *Possible* Worlds

We start with the need for what McKelvey (2001, 2004a) calls *adaptive tension* in the system. New order emerges in a system when the energy differentials within it—adaptive tensions—are of sufficient strength to trigger phase transitions. Emergent events result when this tension—i.e., the energy differentials—exceeds some *critical value*, *R*. A process of new order creation is then initiated and newly emerging patterns begin to appear. For a simple example, consider a pot of water being heated. When the heat produces a temperature above *R*, a phase transition occurs and the water molecules begin moving in a rolling boil—a new structure—instead of remaining stationary and increasing their vibration rate. This basic law in complexity science—dating back to Bénard (1901)—tells us that only tensions above *R* cause emergent events to unfold. We hypothesize that limiting our search to contexts in which the tension exceeds *R*—call these $\boldsymbol{k}$-*contexts*—offers an effective method for filtering out most of the irrelevant dots, links, and possible patterns, leaving only those worth more attention.

Adaptive tension is the motivating force that activates transactions (interactions or links) among "dots". Adaptive tension can sometimes be increased or diminished through pragmatic action as described in Figure 1. The tension has to be high enough to exceed the threshold-activation levels of the relevant dots. If no dot gets activated there cannot be an emergent transaction. Tension, then, separates out relevant transactions from the rest, turning them into candidates for further processing—i.e., they belong to some *possible* world, one that provides some $\boldsymbol{k}$-context in which they have relevance. It follows that, in a $\boldsymbol{k}$-context, if the adaptive tension remains below *R*, the transactions will not form part of the emerging coarse-grained structure of relevant patterns. In the case of the beeper, for example, just going and asking the white middle class female "Is this your (orange) bag?" may be effective in generating the required level of adaptive tension—how might she react? It is unlikely to do so, however, if one simply puts the question at random to other passengers that happen to be in the subway car.

3.2.2 Step 2: Identify Corroborating Vantage Points and *Plausible* Worlds

Having established a pattern's relevance in one of the $\boldsymbol{k}$-contexts—i.e., a possible world—additional corroboration from unbiased perspectives is essential—moving patterns from possible to "*plausible*." We need multiple vantage points, $\boldsymbol{v}$, for this added corroboration. If a possible pattern makes sense from the "situated" perspectives of different agents—i.e., it gets triangulated—then it gains in plausibility.

Corroboration through $\boldsymbol{v}$-vantage points narrows the field of possibly relevant coarse-grained pattern-elements even further, summing over non-corroborated tension-driven transactions to relegate them to the fine-scale structure. It constitutes a second filter that moves us towards coarse-graining and offers two advantages: (1) $\boldsymbol{v}$-vantage points, say three, offer corroboration that a particular emergent transaction—activated by some tension—is relevant and not a random transaction driven by some tension below *R*; (2) $\boldsymbol{v}$-vantage points offer the chance of retrieving a relevant transaction that might have been missed by one agent's $\boldsymbol{k}$-context filter but picked up by another's.

In our beeper example, if another passenger's beeper had fired at the station where the woman boarded the train, if she had been observed to behave suspiciously by a station employee, and if all this had been

corroborated, then, from three distinctive ***v***-vantage points, one would have reason to process her data further. It may be that the first passenger ignored his sensor's signal. No pattern suggested itself given his ***k***-context. But now we have triangulation and the first sensor's signal needs to be retrieved and attended to.

3.2.3 Step 3: Identify Time-based Robustness and *Probable* Worlds

Ashby (1962) observed that chaotic environments do not impose the kind of variety (the degrees of freedom defining complexity) that an organism or an organization can easily respond to. Just as a teapot on a stove requires a continuous source of heat to come to the rolling boil, so our dots, for the most part, only get configured into patterns exhibiting some stability if the adaptive tension is *consistent* over a number of time periods. A tension that comes and goes or changes its nature is chaotic and, thus, cannot consistently instigate the progressive emergence of intelligible patterns over time. Consequently, tracking phenomena across somewhat differentiated (mildly chaotic) time-periods accomplishes two things: (1) Patterns that, are robust enough to persist for some time under adaptive tension and in mildly chaotic conditions—i.e., they remain correlated across sequential time periods—are indeed worth worrying about; and (2) Given $\boldsymbol{d} \times \boldsymbol{h}$ dot-vectors, even after testing for ***k***-contexts and ***v***-vantage points, a large number of transactions could remain that do not lead to an emergence sequence. Further filtering can be achieved by looking for correlated elements that yield patterns across some given number of time periods.

But what constitutes a relevant number of ***t***-time periods? How far, for example, should one work one's way back in time up the beeper trail? Was the adaptive tension that generated the current terrorist concern born shortly before 9/11, or should we go further back? To the Gulf war of 1990 (Burke, 2003)? To the creation of the state of Israel in 1947 (Keppel 2004: Roy 2004)? To the fall of Granada in 1492 (Fletcher 1992)? Alternatively, across how many future time periods should one wait for events to unfold? Waiting too long could allow the emergent pattern to crystallize, with the result that some emergent event such as 9/11 actually occurs. For example, the FBI could be said to have been waiting for the n^{th} time period, which, as it turned out was (obviously) at least one time period past the date of the 9/11 attack. Yet in retrospect we see that the adaptive tension had been building up decades, if not centuries before the event. Indeed, patterns specifically presaging 9/11 were discernible months before the event (Strathern 2003). There is no magic figure to offer for what might be appropriate number of ***t***-time periods. Nevertheless, ***t***-time-period filters will reduce the set of plausible patterns to one of probable patterns by focusing only on those that recur enough over time to yield a frequency count.

To summarize, we can represent our matrix of interlinked dots as a densely connected network that is gradually thinned out by successive filtering—i.e., by a sequence of matrix multiplications that reflect the selective influence of ***k***, ***v***, and ***t***. Such a network can then be made to behave as a *connectionist computational model* in which transactions now look like the correlated firings of neurons.

3.3 COMPUTATIONAL STRATEGIES

In moving from possible to probable worlds via plausible worlds one confronts the problem of computational overload. How does one deal with it? The computational strategies for dealing with such overload will differ in each step. Two such strategies have been established over the past thirty years: Serial and parallel computing. They can be made to work in tandem. Taking each in turn:

Serial Computing

A serial computer (SC) uses a "top-down" process that takes symbolic material as its input, that is, data that have already been processed into information in the form of symbolic patterns. It emits symbolic material as its output after having submitted it to a set of well-defined operations. A sophisticated example of such a process is *structural equation modeling* (SEM) (Kaplan, 2000), illustrated in Figure 4:

1. For those familiar with questionnaire design, our text-message terms correspond to scale items that act as inputs; our pattern elements—part of a pattern-structure—correspond to those variables that make up the theoretical explanatory structure. The dependent variables get "chunked" into prospective "training models" that, when finalized, will constitute inputs for the neural net model. In Figure 4a, the ***k****-contexts* show text messages getting processed and filtered into a number of candidate dependent variables. In Figure 4b, the ***v****-vantage point* further filter these "plausible" patterns, with some now becoming more dominant; Figure 4c shows the filtering effects of the ***t****-time periods* now, generating one or two pattern elements that now become plausible training models for parallel processing by the neural net.

2. In the SEM, text messages are first filtered by the context in which different GNW members find themselves, then by vantage point provided by individual GNW members, and finally over successive time periods. The program starts with specified contexts, such as: terrorists, cults, rebels, pandemic disease indicators, new technological elements, etc. Or, it may find contexts as it analyses messages. As texts are passed through ***k*-, *v*-,** and ***t*-**filters, the emerging candidate patterns become fewer and more focused. Through successive iterations, the SEM provides ever better training patterns to the neural net parallel processor.
3. The objective of the SEM computation is to *reduce candidate patterns* generated by countless possible input data/patterns.

<<< Insert Figure 4 about here >>>

Parallel Computing

A Parallel processor (PP) or connectionist computer uses a "bottom-up" process designed to deal with sub-symbolic material as its inputs. The neural network processor (Hassoun 1995) depicted in Figure 5 is a typical example. By processing this material through a set of "hidden layers," the PP generates plausible patterns as its outputs. Through successive iterations, these eventually coagulate in to probable and therefore actionable hypotheses. In order to find plausible patterns more rapidly, PPs are often given a "training pattern" against which to test those that they generate for goodness-of-fit. The training patterns act as "Bayesian priors" which, though a series of iterations, get gradually updated. Where do such patterns come from? In our case, the SC provides them:

1. Our PP is a neural net that takes a "*Chunked*" output from the SEM as a temporary training pattern. It keeps adjusting its "hidden variables" in an attempt to achieve a goodness-of-fit between the patterns that it generates and those it has received as inputs. Initially it achieves a number of plausible fit and then later a very few probable fit. Its "Bayesian updates" generate the best matches with the training patterns (the *Chunks*) produced by the SEM. Eventually they become identifiable actionable patterns for reaching back into the real world to uncover the terrorist "dots."
2. The objective of the neural net computation is to find the *few relevant terrorist dot-links from among all those* that match the training pattern.

<<< Insert Figure 5 about here >>>

As already suggested, if it is to act as an efficient pattern processor, a PP requires the guidance of training patterns. The difficulty is that training patterns themselves need to get updated dynamically as time progresses in response to new text message inputs—the world does not come to a halt while we try to figure it out! The PP and SC therefore have to work in tandem, as depicted in Figure 6 in order to achieve convergence on some increasingly probable pattern. In our scheme, then, working forward through ***t*-**time periods, the SC *finds the few key patterns from among the trillions of possible ones* that can act as training models for a PP. The latter, working backward in time, and drawing on all the dot-links provided by the GNW, can now *find the few key dots-links that generate the most probable and hence actionable pattern*.

<<< Insert Figure 6 about here >>>

So far, we have described the computational process in the abstract. How is it going to be made to work in practice? Beyond registering a few beeper signals, a purely human system, on its own, would appear to be incapable of rising to the pattern-processing task we have described. Nor, on its own, could our SC and our PP, even working in tandem: the number of "training patterns" required to initiate the computation is itself still far too large, varied, complex, and changeable. No one pattern of interlinked dots could be credibly selected in advance. What we need to do is to combine the remarkable pattern-processing skills of human agents with the data-processing capacities of computers, a hybrid ***socio-computational*** approach involving Global Neighborhood Watchers equipped with detectors, beepers, cell phones, etc., providing inputs—both dots and patterns of dots—to the two computational processing components described above:

- ✓ A *socially distributed network* (SDN) of Neighborhood Watchers, with beepers, cell phones, text-messaging capability, and cell-phone-based photography, each providing "dot-vector" inputs to serial and parallel computing processes.
- ✓ The SDN first provides inputs for a *serial computer* implementing a structural equation model (SEM). This evolves an optimal mix of candidate training patterns that improve over time as the SEM cycles through the ***k***, ***v***, ***t*** corroboration processes and as dot-vector inputs get contextually enriched and updated.
- ✓ The PP is "trained" over time to uncover dot-links comprising the most actionable patterns—in our case patterns generated by a local human interpretation of "dot-vectors" provided by beepers, situational contexts, and other input data. The integration of people, sensors, beepers, and cell phones creates a socio/computational device that harnesses a human being's natural pattern recognition and sense-making abilities to the data processing and transmission capabilities of sensors, beepers, etc., and then to the high-speed data processing capabilities of the NCTC computer.

- ✓ The cycling of the SEM between continually improving contextually enriched text messages, improved temporary training patterns (TTPs), and the re-contacting of GNW members so as to further corroborate what they are seeing and interpreting, eventually results in highly probable training patterns for the PP to search for.
- ✓ The SDN essentially keeps looking for key interlinked dot-vectors, even as its training patterns are being constantly updated by the SEM.
- ✓ These processes result in the timely reduction of trillions of possible patterns down to a manageable number of interlinked dot-vectors—i.e., patterns—and an identification of the generating dots, that is in our example, the terrorists-relevant dots.

4 DETAILING THE SOCIO/COMPUTATIONAL PROCESS USING A TEST SCENARIO

By passing them through ***k***, ***v***, and ***t*** filters, the SEM cycles the continuously upgraded inputs being received (beeper signals, text messages, photos, conversations, etc.) in order to help the PP to select out the key links (and dots) from among the trillions of possible patterns—i.e., patterns that have coherence, can be recognized, and result in actionable dot connections. A further development of our subway scenario will show how this works. To save space we use a number of acronyms. They are defined in the box below:

Glossary of Acronyms

GNW	Global Neighborhood Watch	SDN	Socially distributed network
NCTC	National Counter-Terrorism Center	SEM	Structural equation model
PP	Parallel processor	STMs	Short text messages
SC	Serial computer	TTPs	Temporary training pattern

4.1 STAGE ONE: k-CONTEXTS AND POSSIBLE WORLDS

The city of Washington is in *Condition Red*. "Chatter" from here and there around the world has accumulated recently suggesting that a possible terrorist action may be aimed somewhere in the city, involving either the explosion of a dirty nuclear bomb, or an anthrax or sarin-gas attack. No one knows which. There is some reason to believe that some of the chatter may also be purposefully misleading. A number of years back, however, throughout the US and in a number of other countries, GNW organizations were created, each member carrying a bio-chemical/radiation detection device, a beeper, and a cell phone with photo capabilities. Washington is a large city and the Washington chapter of GNW has several thousand members anyone of whom can also be contacted via the network.

The subway train on which you are riding is now about ten minutes from Union Station. Riding with you on the train are other GNW members on their way to work, going shopping, etc. Triggered by a chemical sensor, the beeper that you are carrying has just gone off and sent a signal to the NCTC Computer that stores it temporarily. A couple stops later, Gillian and Veronica—the former, a GNW member—enter the other end of your car whereupon Gillian's beeper goes off (see Introduction).

When two or more beepers—i.e., dots—go off at the same time, a correlation is established. Does it suggest a possible pattern?

- ✓ Most of the time, the beepers go off here and there in a random fashion. In most instances, no inter-dot connections are implied by such false positives.
- ✓ If the triggering of two beepers more or less coincides—say caused by the action of a chemical sensor—then the signals automatically get registered in the Computer. An agent located at the NCTC can then communicate directly via short text messaging (STM) with the GNW members wearing the beepers through their cell phones or it can alert a higher-level agent, who then "conference" calls both of them.
- ✓ At this point the exploration of possible patterns is initiated. The GNW members in the train look around the cars they are in. One sees a man and woman with a baby carriage. The man has a "Mediterranean" look. The NCTC agent asks the GNW members to get nearer; the beeper signals from their detectors get stronger.
- ✓ The signals emitted by the various beepers all get registered in the NCTC's Computer. These messages could be automatic, having been triggered by the beepers, or they could be human-activated messages transmitted via a cell phone. The latter could be verbal, with verbal word recognition, STMs with computational text recognition, or photos sent to a graphic recognition processor in the Computer.

Only *possible* patterns (some subset of possibly connected input signals) are generated at this stage. In this case, two beepers (dots) firing in a subway car initiate a pattern generation process between them. This instance of correlation between firing beepers need not be a purely local affair. It could extend to

international locations such as a garden in Perak (Malaysia), a Mosque in London, Singapore's Changi Airport, a school playground in Karachi, or a casino in Macau. No coherent story—i.e., pattern—can yet account for these correlated beeper firings. That is, the pattern cannot yet be *recognized*. A huge number of possible candidate input patterns, therefore, are in need of processing. The vast majority of them will get eliminated when passed through ***k***-context filters.

When it can correlate over space and time—if only weakly—the beeper signals that it receives, the SC initiates a top-down process. The next step is to automatically call back the relevant GNW members and get them to briefly describe the different *contexts* in which the beeper signals were emitted. Pattern recognition now becomes both a *local* and a *human* affair. GNW members can communicate verbally with an NCTC agent or via STMs to the SEM—located in the Computer. The contextual information received at the NCTC's Computer is then associated with keywords relevant to the inputs being received.

4.2 STAGE TWO: v-VANTAGE POINTS AND PLAUSIBLE WORLDS

Beepers going off at random here and there can, for the most part, be ignored. But if several go off in some correlated fashion, then keywords can be derived from the different vantage points of GNW members to guide further data processing and interpretation. In this way patterns become more plausible:

- ✓ The NCTC agent calls the GNW members whose beepers initiated the process and asks them to describe in some detail what they see. In this specific case, the agent then alerts other GNW members who happen to be either on the subway train or in the vicinity of its destinations—Union Station and the other stations the train is due to stop at—to become more alert and to converge toward these stations, etc. Given some corroborating overlap in the description of GNW members located at ***v***-vantage points, candidate input *pattern elements* are identified and processed. These are evolved in an iterative fashion through the continuing interaction of the NCTC agent and the SC on the one hand with the GNW members, the SDN, and PP on the other.
- ✓ Upon receiving beeper signals, the SEM is programmed to request text messages from GNW members using as many standard code words as possible—i.e., Central Asian, young male, bearded, strange expression, box, suitcase, laptop, shopping bag, stroller, etc. The SEM is programmed, to rank these words against Zipf's (1949) power law ranking (which starts with: *the, of, and, to, I, or, say, really, quality….*)—looking for frequently used words that are inconsistent with the power law ranking of frequency of use.
- ✓ Guided by the NCTC agent, the SEM keeps cycling through all incoming messages trying to tease out plausible connections—*pattern elements*—between them (a process depicted in Figure 6). The SEM then suggests suitable terms or labels, to "help" GNW members communicate the pattern(s) they see. Of course, the SEM should not be so "helpful" as to lead members to "see" non-existent patterns. As the SEM's "experience" accumulates it gets better at suggesting appropriate terms—those learnt in earlier training exercises and that GNW members themselves would want to use. The socio/computational system —comprising GNW members, SDN, NCTC agents, SC, PP, and SEM—would earlier have participated in various pattern-generating training exercises.

As pattern elements emerge, get corroborated across sensors, GNW beepers, and v-vantage points—i.e., as dots get "filled in" and links are found—candidate patterns begin to take shape. As pattern elements accumulate and get further defined, the nascent patterns gain in plausibility.

4.3 STAGE THREE: t-TIME PERIODS AND PROBABLE WORLDS

When two (or more) GNW inputs generate correlated *keyword-and-context* messages across *t*-time periods, the recurrence justifies the formulation of probabilistically derived expectations. As more GNW members approach the vicinity of the subway train, the number of possible incoming inputs could well increase far beyond an NCTC agent's capacity to respond. Here is where SEM becomes especially useful. Increasingly plausible patterns recurring across *t*-time periods now produce temporary training patterns as inputs to the PP:

- ✓ At some point the SC runs "*block model*" software[2] to more efficiently abstract out training patterns from the inputs. Needless to say, the SEM needs some minimum number of inputs to get going. Via the abstracting process, redundant pattern elements begin to coagulate into *candidate* pattern elements that get integrated into meaningful "wholes." These the SEM consolidates into trial sets of pattern elements, *Chunks*, that are entered into the PP to serve as *temporary training patterns* (TTPs) for it.
- ✓ Each TTP—and there could be many of these instead of just one as with conventional neural net models—now acts as a template against which further inputs can be compared. An emergent pattern would grow in strength to the extent that its constituent elements matched the SDN-generated patterns that now define a TTP. The larger the number of spatio-temporal events that match a given TTP, the more probable it becomes.
- ✓ For each initial set of two (or more) inputs that end in a TTP, the recursive cycling ends when the most successively refined, and thus, most probable TTPs are fed back to the SDN for final checking. At this stage we have a TTP built up from the

various inputs emanating from interactions with *situated* GNW members. TTPs thus consist of *probable patterns*—alternative configurations of dots and links enjoying a high degree of corroboration.

Stage Three ends with the emergence of a limited number of sequentially corroborated TTPs.

4.4 STAGE FOUR: PATTERN MATCHING IN ACTUAL WORLDS

TTPs are testable hypotheses. Sooner or later, nature will shows its hand and tests them. By this time, in our scenario we have potentially hundreds of GNW members converging on subway stations along the train's route. They have generated a number of TTPs for the SEM to process as probable scenarios. Many of these diverge, but some converge to produce *Chunks*.

- ✓ While cycling through and refining TTPs, the SDN compares them and looks for *Chunking* opportunities.
- ✓ *Chunks* are defined by standard bio-taxonomic method: No single term is essential; most—but not all—are included.
- ✓ TTPs, generate chunks in two ways:
  - From overlap with or corroboration by other TTPs. They gain credibility from redundancy and are now reduced in number. The block model abstracts and synthesizes multiple patterns into Chunks and the SEM now sends these out to GNW members for further corroboration/confirmation.
  - From the integration of *new* incoming recycled STMs providing further corroboration that individual links between dots constitute valid pattern elements.
- ✓ TTPs now either: (1) merge into a single Chunk; (2) are cast away as errors; (3) are maintained as alternative hypotheses.

Here is where SEM "logic" appears. SEMs "compute"—that is, they pick selected pattern elements as inputs and modify their interrelationship thus producing structural variables—so as to ultimately find the selection of data inputs and patterns that most accurately define the dependent variable (*Chunk*) to be predicted. Beeper signals, STMs, photos, and taped conversations are the *pattern elements* to be input. The actual *patterns* play the role of the "structure" in the SEM. The inter-relationship of their components is constantly changing, just as independent and moderating variables change position in SEM modeling as correlations strengthen or weaken. The *Chunks* play the role of the dependent variable(s)—as pattern elements are added or deleted and as structural elements (the TTPs) are rearranged into trial patterns, the *Chunks* become better defined. The SEM program cycles through the various pattern elements, TTPs, and *Chunk* constructions until one or more coherent *Chunks* emerge. First, following block model theory, the challenge is to represent the commonalities among the TTPs with the fewest *Chunks*. This results from abstracting out the more frequently reoccurring links across the population of patterns—the pattern(s) built from the dominant abstractions (those occurring in the most patterns) become *Chunks* (Wasserman & Faust 1994). Second, in structural equation modeling (Kaplan 2000), the challenge is to maximize the correlation of independent variables (the TTPs arising from the STM terms) with emerging *Chunks*. Collapsing similar TTPs into a single *Chunk* accomplishes both.

Chunks, the output of Stage Four, constitute *inferences to the best explanation* (Sklar 1995)—hypothetical patterns systematically evolved to train the PP in what to look for as it sifts through the vast number of dot-vectors supplied by all of the input signals.

4.5 THE LEARNING CHALLENGE

As Ashby's law implies, pattern generation needs to progress at a faster rate than the rate at which emergent events actually unfold. This requires expeditious learning. First, as GNW members become better trained to perform in SDNs—through the timely identification and accurate description of context—the speed of pattern corroboration, convergence, and validation achieved by the SC improves. There will now be less volatility of pattern elements as new inputs come in. Indeed, the reduction in volatility offers a good proxy measure of corroboration, convergence, and validation at work. Second, as the SC, through periodic practice drills, gets better at "interpreting" and "assisting" GNW members, it will end up with fewer but more valid TTPs. These now become an increasingly sparse SEM "structure" with fewer actual patterns (SEM variables), and stronger connections between pattern elements. Finally, as the SEM structure improves, the number of seemingly different *Chunks* decreases. These can now be tested against an unfolding reality—do they suggest a false alarm or do they foreshadow an attack on Union Station? Or some alternative target?

5 ORGANIZATIONAL IMPLICATIONS

5.1 THE CHALLENGE OF INTELLIGENCE ASSESSMENT

What are the organizational implications of the above analysis for intelligence services? As we saw, the quality of intelligence assessment was identified as a major problem in post 9/11 analyses. Who does it? How is it managed? Returning to our scenario, should the assessment of what is going on—i.e., pattern recognition—be centralized in some distant office, or should it be decentralized to locally "situated" agents? What if one was dealing with a train station in Peshawar (Pakistan) rather than Washington? And how fault-tolerant can such pattern recognition be?

The new NCTC aims to bring all the analysts and spies working on counter-terrorism under a single authority. The hope is that the new reforms will help to "fill in the dots" more rapidly and accurately next time by promoting a better sharing of information among the experts—so as to find more dots more quickly. Unfortunately, this proposal leaves intelligence assessment—i.e., *pattern recognition*—essentially undisturbed at the top of the hierarchy, and thus subject to the same problems of data overload and of filtering biases that we discussed earlier.

Furthermore, putting the intelligence services under a single authority does nothing to remove the silo mentality that operates both within and between these services. While the 9/11 Commission's *Final Report* describes the CIA, the FBI, and more than thirteen other intelligence units as "cast-iron stovepipes" at the agency-level (p. 403), silos also exist *within* these agencies. The FBI, for example, already had one boss, and yet people from different parts of the organization—its internal silos—didn't talk to each other (Posner 2003). In effect, the "stronger management" that the Commission calls for to unify all the intelligence-gathering agencies (p. 411), will simply replicate at a higher level the existing silo-creating management approaches existing at the CIA, NSA, FBI, DOD. And as we now know, the hierarchical structure of these agencies exacerbates rather than ameliorates the silo problem.

By connecting GNW members directly to a new pattern-processing technology, and by taking advantage of thousands of "open source" information opportunities, our "distributed" socio/computational approach gets around silo thinking. How? A hierarchical organization deploys information filters in such a way that information is extracted from data collected at the base and is then passed up to the next hierarchical level. That is, dots (data) collected at the base get "joined" or linked (information) by intelligence analysts in the middle of the hierarchy before being "assessed"—i.e., wrought into patterns (actionable knowledge)—at the top. The finer the lower level filters the slimmer and steeper the hierarchy, as depicted in stylized form in Figure 7a. By contrast, filtering for a distributed approach would produce an organizational structure looking more like Figure 7b. Whereas in Figure 7a, agents at the bottom eliminate dots at a rapid rate in order to limit the data processing load of those higher up the hierarchy, in Figure 7b, lower-level agents deploy a limited number of dots to generate a larger number of links and a vastly larger number of pattern elements requiring higher-level processing.

<<< Insert Figure 7 about here >>>

The difference between the two figures reflects fundamentally different data processing strategies. In Figure 7a, relevant information is presumed to reside primarily in the mean of a data distribution. Under assumptions of a normal distribution and independent events, the variance is mostly treated as an error term, i.e., noise that has to be got rid of as quickly as possible. In Figure 7b, by contrast, relevant information is presumed to reside primarily in the variance—independence of events and a normal distribution can no longer be assumed. This is the source of the combinatorial explosion that generates so many patterns for processing further up the hierarchy. In the 7b case a large amount of information-bearing data has to be moved up the hierarchy before anything useful can be extracted from it. Focusing on means is what intelligence services were acculturated to do during the Cold War when the enemy and its possible range of behaviors were relatively well known and it was relatively easy to tell what was data and what was noise. As Thomas Kean (2004), the Chairman of the 9/11 Commission observes, however, the national security bureaucracy is still stuck in the Cold War time warp. In a post-cold war world of asymmetric threats in which small causes can have disproportionately large effects, intelligence services must learn to focus on information-bearing variances as much as on means (Gleick 1987, Andriani & McKelvey 2004).

Clearly the volume of data processing implied by Figure 7b is many times larger than that implied by Figure 7a. *Too large, we believe, to be effectively processed by any traditional organization or even a limited number of these such as the CIA, the FBI, Homeland Security, etc.* The implication of Figure 7c—a superimposition of Figures 7a and 7b—is that the intelligence process needs to spread outside existing hierarchical organizations and take advantage of *socially distributed computational* methods (Prietula, Carley, & Gasser 1998). Furthermore, we now need to distinguish between traditional hierarchical *data* processing and socially distributed *pattern* processing. Whereas the former type of processing still leaves intelligence assessment—i.e., pattern processing—exclusively in the hands of a few overloaded, higher-level "experts," the latter type distributes pattern processing activities across a much wider community, replacing the hierarchically located—and thus silo-resident—"experts" with the concept of a Global Neighborhood Watch (GNW) composed of ubiquitously located ordinary citizens. Assessment does not thereby disappear, but it now operates across different levels: at the lower levels of the hierarchy, meaningful patterns are intentionally *generated* and corroborated for subsequent selection and processing; at higher levels—possibly that of the NCTC—the most promising of these are actually put into play and fed back to the lower levels in an interactive fashion. The implication of our scheme—and what our scenario builds upon—is that the NCTC needs to involve many more stakeholders than existing intelligence services. But to avoid information overload it also needs a way of organizing that is designed to take advantage of our proposed socio/computational technology.

5.2 WHO IS MY NEIGHBOR?

A culture that supports distributed pattern processing differs significantly from one that supports sequential data processing alone. In our approach, GNW members become "*coalitions of the watchful.*" Intelligence gathering, checking, corroborating, and contextually enriching, now become the business of citizens everywhere, as well as of "silo-filled" government agencies. Our GNW concept can be configured differently for different missions:

- ✓ Watch organizations can spread across the entire United States and beyond—i.e., to US citizens abroad, to green-card carriers who travel, and, via foreign governments, to foreign nationals;
- ✓ GNWs may be put through training exercises, be nationally cognizant, be focused on local hot-spots, be relatively easily reequipped, have equipment more easily maintained, upgraded, etc.;
- ✓ GNWs can be rapidly activated and deployed within and around metropolitan areas. In their capacity as employees, they can be moved in and out of specific sites such as nuclear power plants, container unloading docks, airports, subways, government buildings, and vulnerable office buildings;

The pre-9/11 FBI story (Posner 2003, Anonymous 2004) illustrates the importance of having numerous "*agents*" actively exploring the vast number of terrorist-relevant dot-vectors and then sharing their findings. Graham (2004: 243–244) emphasizes the need to "…expand the number and orientation of *voices* that contribute to the intelligence process." He also says "the intelligence community…needs to be more amenable to the use of intelligence collected from *open sources*" (our italics). We know from recent experience that existing Neighborhood Watch organizations, the ultimate "open source," contribute significantly to crime reduction. Could similar citizen-based organizations help to contain the terrorist threat?

The GNW concept, however, is not without drawbacks. First, it requires ordinary citizens to take on the role of secret agents and to snoop on other citizens neighbors—in churches, mosques, schools, supermarkets, work places, transportation systems, and the like. Thus, each member of a GNW scheme spies on those around him/her. Since Neighborhood Watches exist in many urban neighborhoods, citizens are obviously willing to play this role to further protect their children and neighborhoods.

Second, diverse ethnic, religious, work, and educational groups could come under pressure to help make the U.S. or their city safer by joining the local GNW organization—if only to demonstrate "whose side they are on"—thus destroying the concept's essentially voluntary character. Why is it important to maintain the scheme's voluntary character? The very concept of a "neighborhood watch" scheme, invites the question: "*Who is my neighbor?*" In a rapidly globalizing world, the answer must be: "*The stranger that I have never met and who may be affected by my action*." To succeed, GNW has to become part of the struggle for "hearts and minds" of citizens everywhere. As the 9/11 Commission's Vice Chairman, Lee Hamilton

(2004), pointed out, military action and heightened security will never be enough. He argues that the United States should promote "an agenda of opportunity" in impoverished countries and join "the battle of ideas" so that those regions do not become incubators of future terrorists—as has Iraq. At present too many people outside the West—and, as Gilles Keppel (2004) has recently stressed, a good many in the West itself—harbor more neighborly feelings towards Osama Bin Laden than towards those who live in the West and who espouse its values. In our scenario, we present GNW as a mostly local, domestic affair. Yet for the concept to really deliver, one needs to devise *neighbor-generating policies* across national and cultural boundaries. This, of course, is as much the province of foreign policy as of intelligence gathering

6 CONCLUSION

We have drawn upon Ashby's *Law of Requisite Variety/Complexity* (1956) to analyze the nature of the security challenges that nation states face in the 21st century. Under competitive conditions, rapid pattern recognition becomes a weapon in a cognitive arms race between adversaries—requisite complexity calls for the generation of adaptive responses in a timely fashion. We have shown that effective counter-terrorism requires a lot more than just "filling in the dots." Given the trillions of possible patterns a few dots give rise to, the cognitive challenge of quickly reducing the vast complexity of externally emergent patterns far exceeds the organizational capacities of a few government agencies, no matter how well equipped these might be. As we have shown, in the real world the trade off between generating data and generating meaningful patterns is time-constrained. The need for *rapid* pattern recognition may thus set a limit to the amount of data that can usefully be collected and processed. This is a new problem that traditional technologies and ways of organizing were never designed to cope with. The policy implications of the above are that we need:

- ✓ Global Neighborhood Watch organizations offering "distributed" pattern processing and corroboration activities—made possible by ubiquitous and mobile human agents equipped with portable detectors, beepers and cell phones, and communicating via STMs. These activities provide a basis for moving methodically from possible to probable knowledge via plausible knowledge.
- ✓ A centralized socio-computational technology that:
  - Uses structural equation (with block) modeling to reduce trillions of candidate patterns down to an actionable few. These then serve as inputs to parallel processors in the form of successively updated training patterns;
  - Uses neural net parallel processing, driven by increasingly accurate training patterns, to focus searches through corroborated open-source information for relevant dot-links existing in the world as input data—especially those that foreshadow some imminent security threat.
- ✓ Global Neighborhood Watch schemes built from local communities throughout the world and comprising a diversity of ethnic, religious, international, and other groups whose "watching" can be selectively activated when conditions warrant [3].

Ours is an optimistic paper.

As far back as 1945, in a classic paper entitled "The Use of Knowledge in Society," Friedrich Hayek showed how socially distributed processing could help citizens and organizations cope with a complex and fast-moving, nonlinear world. Even then, such parallel processing strategies were to be found in markets. In matters of intelligence, however, government agencies continue to operate exclusively through serial—and thus, hierarchical—models of computation and organization (CIA, FBI, etc). The proposed solution of the 9/11 Commission—to create a new intelligence Tsar—simply builds onto this traditional organizing strategy. The complexity of the security challenge suggests that these traditional approaches now need to be complemented by modern socio/computational data processing methods that integrate silicon- and carbon-based agents in novel ways.

We have demonstrated that "joining the dots," whilst *a* problem, is not *the* problem. Finding the computational capability to process and corroborate trillions of possible patterns is the problem. New open-source pattern processing technology will be needed to address it. Yet, since computational strategies are only effective if the underlying organizational capabilities required to implement them are in place, if one changes computational strategy, then, by implication, one needs to develop new organizational capabilities as well. Our proposed solution integrates new computational processes with new organizational ones. Taken together, they suggest that significant organizational and cultural challenges loom if US intelligence agencies are ever to efficaciously anticipate and counteract modern terrorist threats.

ENDNOTES

[1] Alternatively or in addition, detectors could be permanently installed in key locations—like subway cars, train depots, airports, nuclear facilities, container unloading areas, government buildings, etc.; the portable detectors would then corroborate and contextually enrich signals from the stationary detectors.

[2] Block model programs come out of network sociology. They abstract commonalities across multiple networks (Wasserman & Faust 1994).

[3] Clearly, intelligence services will be able to avail themselves of the socio-computational technology that we are describing here for their own purposes. How strongly they should then be integrated with the GNW membership remains an open and highly political question.

REFERENCES

9/11 Commission (2004). *Final Report of the National Commission on Terrorist Attacks upon the United States*. New York: Norton.

Andriani, P. & McKelvey, B. (2004). "Power Law Phenomena in Organizations: Redirecting Organization Science." Working paper, Durham Business School, Durham, UK.

Anonymous (2004). *Imperial Hubris: Why the West Is Losing the War on Terror*. Washington, DC: Brassey's.

Arrow, K. (1984). "Information and Economic Behavior." In *The Economics of Information: Collected Papers of Kenneth J. Arrow*: 136–152. Cambridge, MA: Belknap/ Harvard University Press

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.

Ashby, W. R. (1962). "Principles of the Self-organizing System." In H. von Foerster & G. W. Zopf (eds.), *Principles of Self-Organization*: 255–278. New York: Pergamon.

Bénard, H. (1901). "Les Tourbillons Cellulaires dans une Nappe Liquide Transportant de la Chaleur par Convection en Régime Permanent." *Annales de Chimie et de Physique*, 23: 62–144.

Binmore, K. (1994). *Game Theory and the Social Contract, Volume 1: Playing Fair*. Cambridge MA: MIT Press

Boisot, M. (1998). *Knowledge Assets: Securing Competitive Advantage in the Information Economy*. Oxford: Oxford University Press.

Boisot, M., & Child, J. (1999). "Organizations as Adaptive Systems in Complex Environments: The Case of China", *Organization Science*, 10: 237–252.

Boisot, M., & Canals, A., (2004). "Data, Information, and Knowledge: Have We Got It Right?" *Journal of Evolutionary Economics*, 14: 43–67.

Boisot, M., & Macmillan, I. (2004). "Crossing Epistemological Boundaries: Managerial and Entrepreneurial Approaches to Knowledge Management." *Long Range Planning*, 37/6.

BonJour, L. (1985). "The Elements of Coherentism." In *The Structure of Empirical Knowledge*. Cambridge, UK: Cambridge University Press.

Burke, J., (2003). *Al-Qaeda: Casting a Shadow of Terror*, London: I. B.Tauris

Duhem, P. (1914). *The Aim and Structure of Physical Theory*. Paris: Marcel Riviere et Cie.

Financial Times (2004). "The human factor: 'All is not well in clandestine intelligence collection.'" (July 7th): 15.

Fletcher, R., (1992). *Moorish Spain*. London: Weidenfeld and Nicholson.

Gell-Mann, M. (1994). *The Quark and the Jaguar*. Freeman, New York.

Gleick, J. (1987). *Chaos: Making a New Science*. New York: Penguin.

Graham, B. (with J. Nussbaum) (2003). *Intelligence Matters: The CIA, the FBI, Saudi Arabia, and the Failure of America's War on Terror*. New York: Random House.

Halevy, E. (2004). "In Defence of the Intelligence Services." *The Economist*, 372(July 31): 21–23.

Hamilton, L. (2004). "A 'failure' to protect *the* U.S." *The International Herald Tribune*, July 23rd, pp. 1 & 4.

Hanson, N. R. (1958). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*, Cambridge: Cambridge University Press .

Hassoun, M. H. (1995). *Fundamentals of Artificial Neural Networks*. Cambridge, MA: MIT (Bradford) Press.

Hayek, F. (1945). "The Use of Knowledge in Society." *American Economic Review*, 35(4): 519–530.

Horwich, P. (1995). "Truth." In R. Audi (Ed), *The Cambridge Dictionary of Philosophy*: 812–813. Cambridge, UK: Cambridge University Press.

Janis, I. L. (1972). *Victims of Group Think*. Boston, MA: Houghton Mifflin.

Kaplan, D. (2000). *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: SAGE.

Kauffman, S. A. (1993). *The Origins of Order: Self-organization and Selection in Evolution*. New York: Oxford University Press.

Kean, T. (2004). "A 'failure' to protect the U.S." *International Herald Tribune*, July 23rd, pp. 3–4.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

McKelvey, B. (2001). "Energizing Order-Creating Networks of Distributed Intelligence." *International Journal of Innovation Management*, 5: 181–212.

McKelvey, B. (2003). “Emergent Order in Firms: Complexity Science vs. the Entanglement Trap.” In E. Mitleton-Kelly (Ed.), *Complex Systems and Evolutionary Perspectives on Organizations*, 99–125. Amsterdam, NL: Elsevier Science.

McKelvey, B., & Boisot, M. (2003). “Redefining Strategic Foresight: “*Fast*” and “*Far*” Sight via Complexity Science.” Presented at the INSEAD Conference on *Expanding Perspectives on Strategy Processes,* August 24–26, 2003, Fontainebleau, France.

McKelvey, B. (2004a). “MicroStrategy from MacroLeadership: Distributed Intelligence via New Science.” In A. Y. Lewin and H. Volberda (Eds.), *Mobilizing the Self-Renewing Organization*. New York: Palgrave/Macmillan.

McKelvey, B. (2004b). “Toward a 0^{th} Law of Thermodynamics: Order-Creation Complexity Dynamics from Physics & Biology to Bioeconomics.” *Journal of Bioeconomics*, 6: 65–96.

Popper, K. (1972) *Objective Knowledge: An Evolutionary Approach*, Oxford: Oxford University Press

Posner, G. (2003). *Why America Slept: The Failure to Prevent 9/11*. New York: Random House.

Prietula, M. Carley, K., & Gasser, L. (1998). *Simulating Organizations: Computational Models of Institutions and Groups*. Cambridge, MA: M.I.T. Press.

Omnès, R. (1999). *Understanding Quantum Mechanics*. Princeton, NJ: Princeton University Press.

Quine, W. V. (1969). “Naturalized Epistemology.” In *Ontological Relativity and Other Essays*: 69–90. New York: Columbia University Press.

Sklar, L. (1995). “Philosophy of Science.” In R. Audi (Ed), *The Cambridge Dictionary of Philosophy*: 611–615. Cambridge, UK: Cambridge University Press.

Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). “Threat-rigidity Effects in Organizational Behavior.” *Administrative Science Quarterly*, 26: 501–524.

Strathern, M. (2003). “The Symmetries and Redundancies of Terror—Notes.” Working paper, Cranfield School of Management, Cranfield University, MK43 7HX, UK.

Tarski, A. (1956). In *Logic, Semantics, Metamathematics* (tran. J. H. Woodger) (2nd ed.): 152–278. Oxford, UK: Clarendon.

Thagard. P. (2000). *Coherence in Thought and Action*. Cambridge, Mass: MIT Press.

The Economist (2003). “America Needs More Spies.” 368(July 12): 30–31.

The Economist (2004). “No Blame, No Pain.” 372(July 31): 30.

The International Herald Tribune (2004). “A ‘failure’ to protect U.S.” (July 23): 1, 4.

von Uexküll, J. (1921). *Umwelt und Innenwelt der Tiere, 2, werm. und verb. Aufl.* Springer,

Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Cambridge, MA: Addison-Wesley.

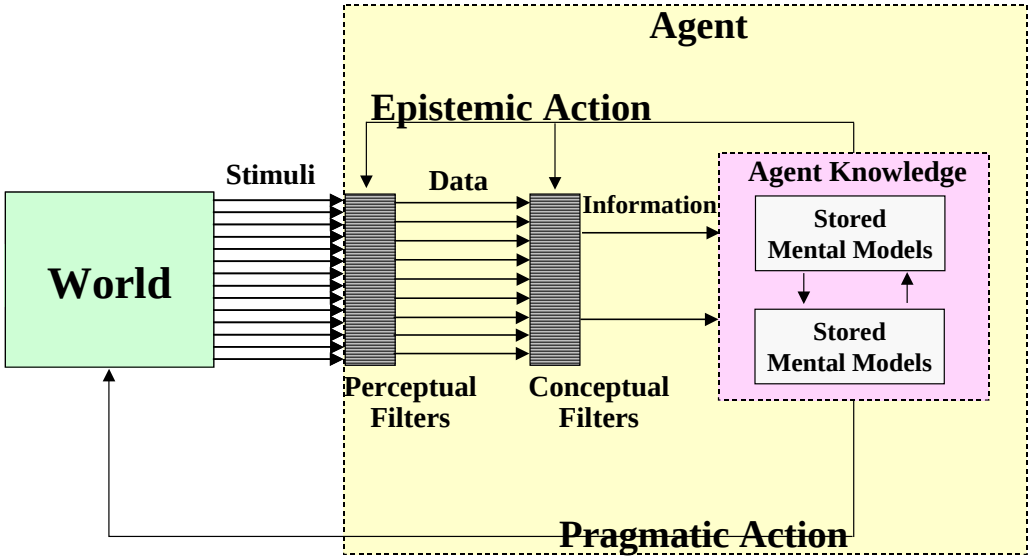


Figure 1: The Agent-in-the-World

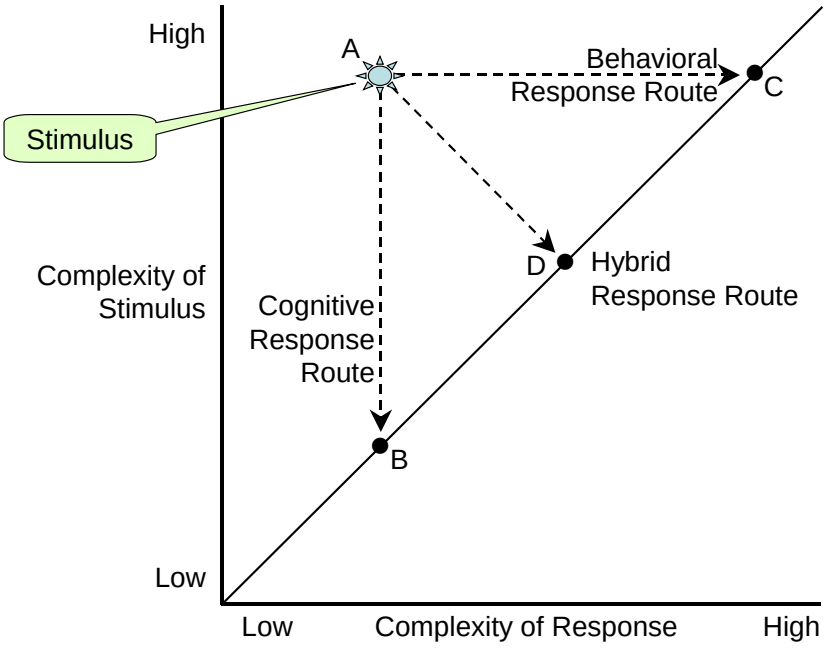


Figure 2: An Adaptation of the Ashby Space

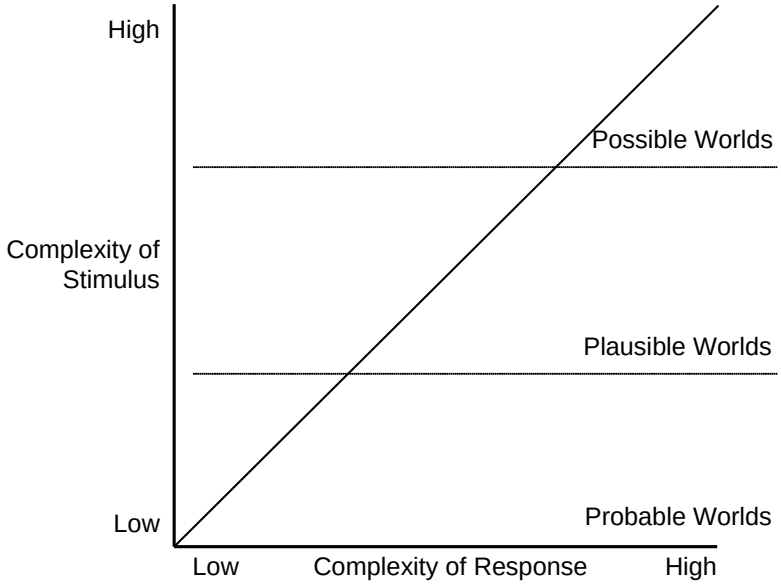


Figure 3: Phase Transitions in the Ashby Space

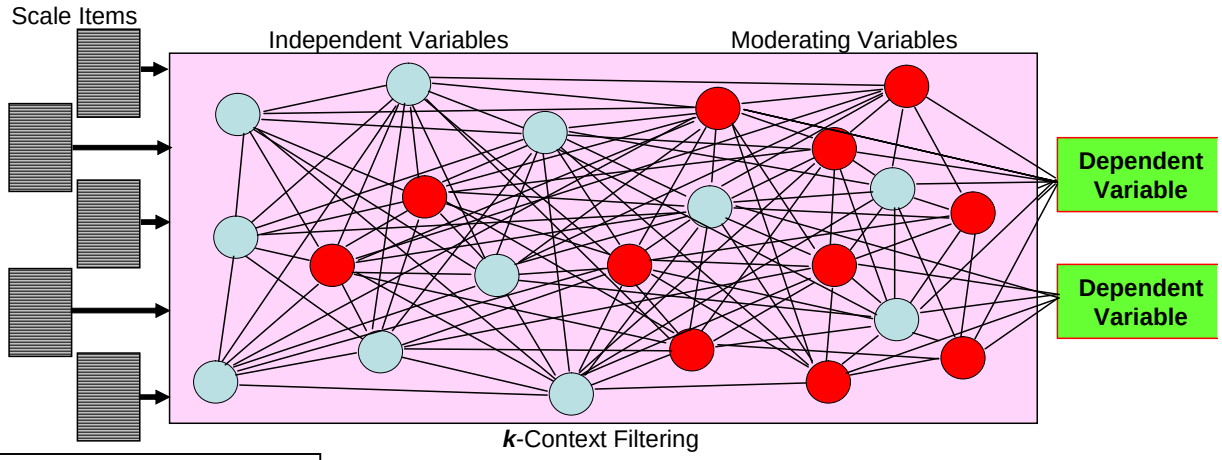


Figure 4a: *k*-Contexts

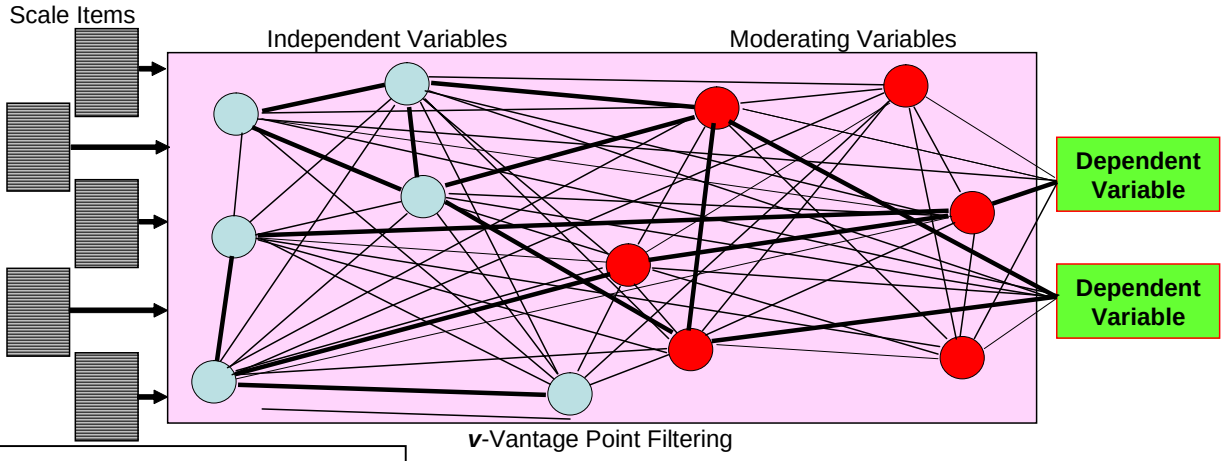


Figure 4b: *v*-Vantage Points

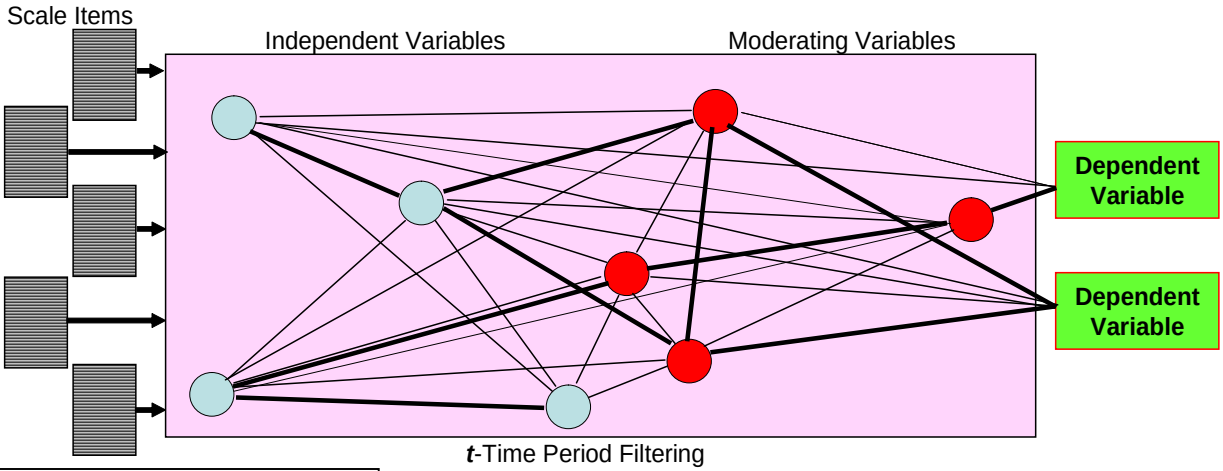


Figure 4c: *t*- Time Periods

Figure 4: Pattern Filtering via Serial Computing

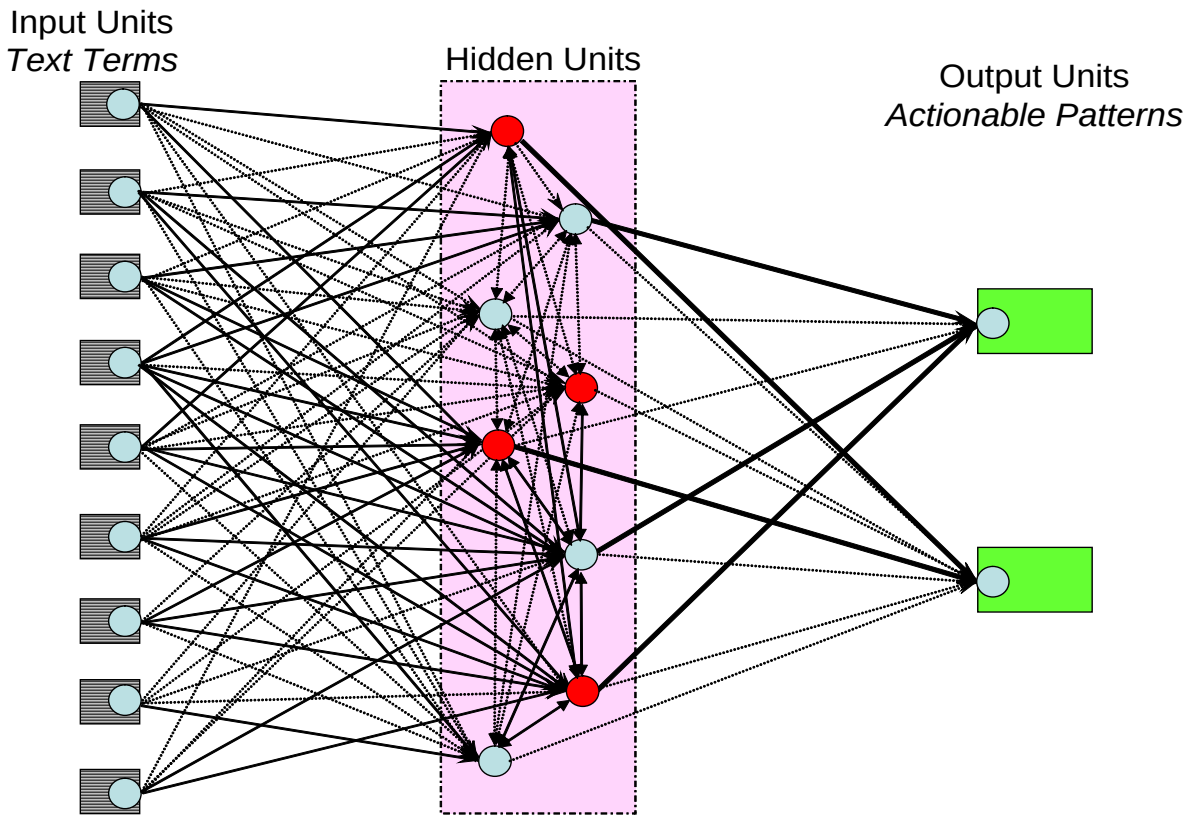


Figure 5: Parallel Computing

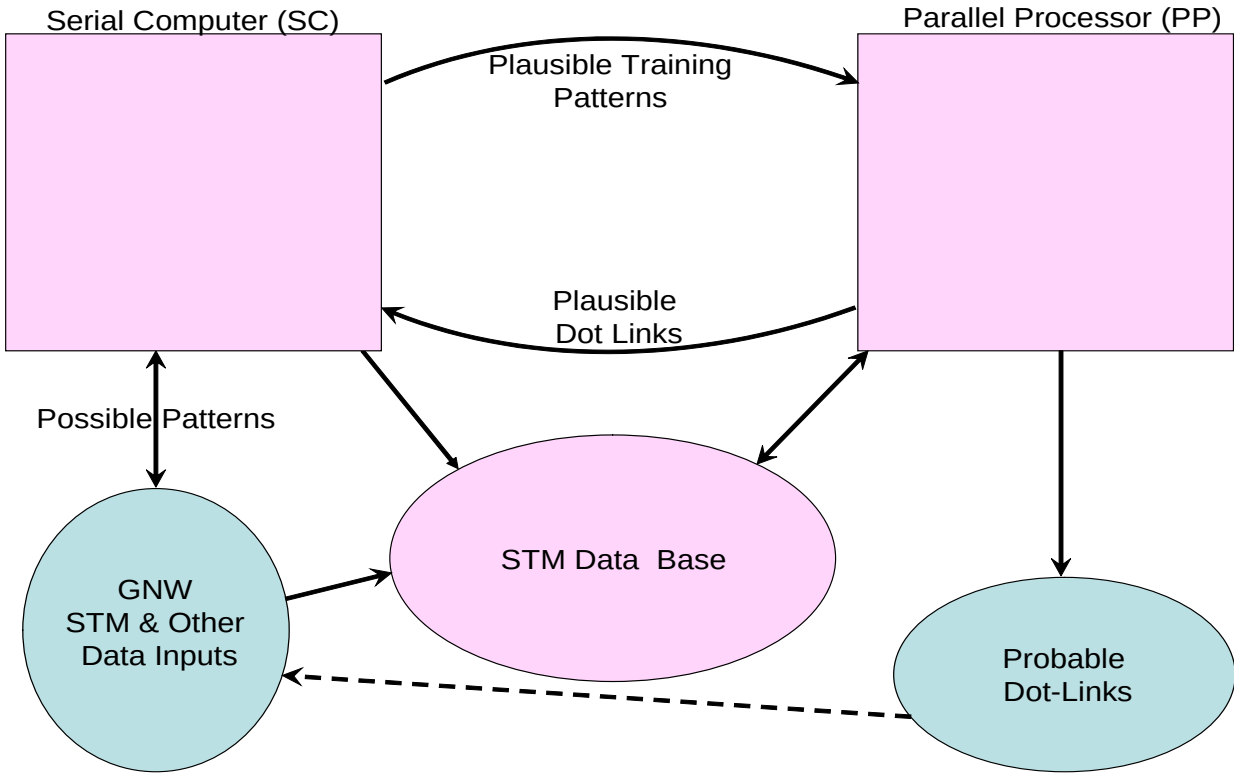


Figure 6: Serial & Parallel Processing Combined

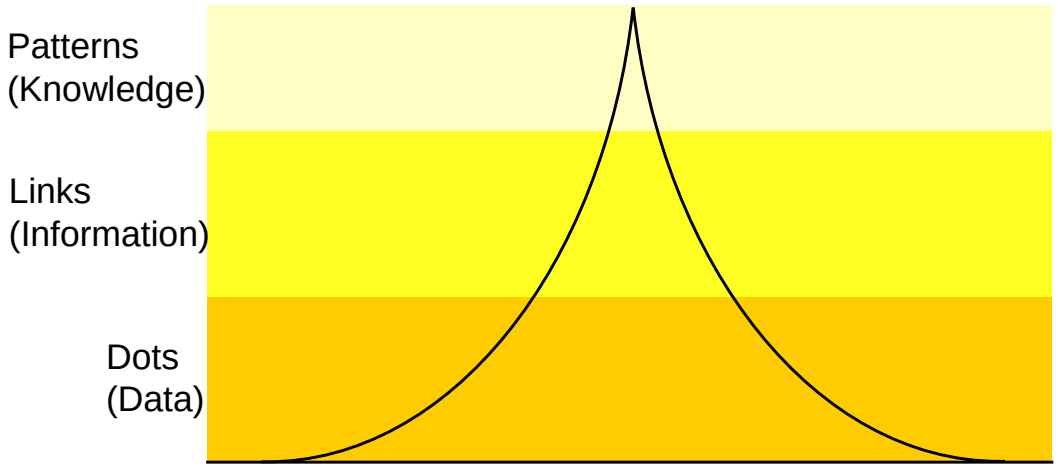


Figure 7a: Traditional Data Processing Hierarchy

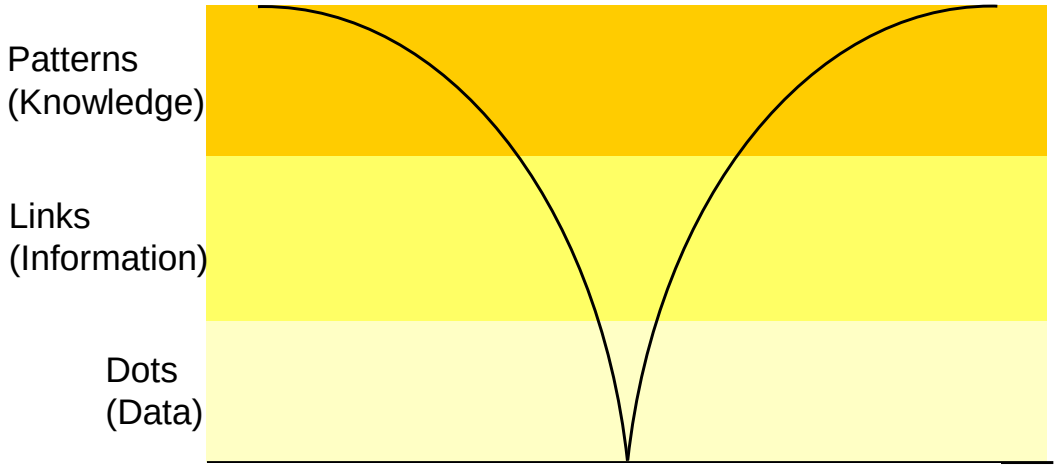


Figure 7b: Pattern Processing Hierarchy

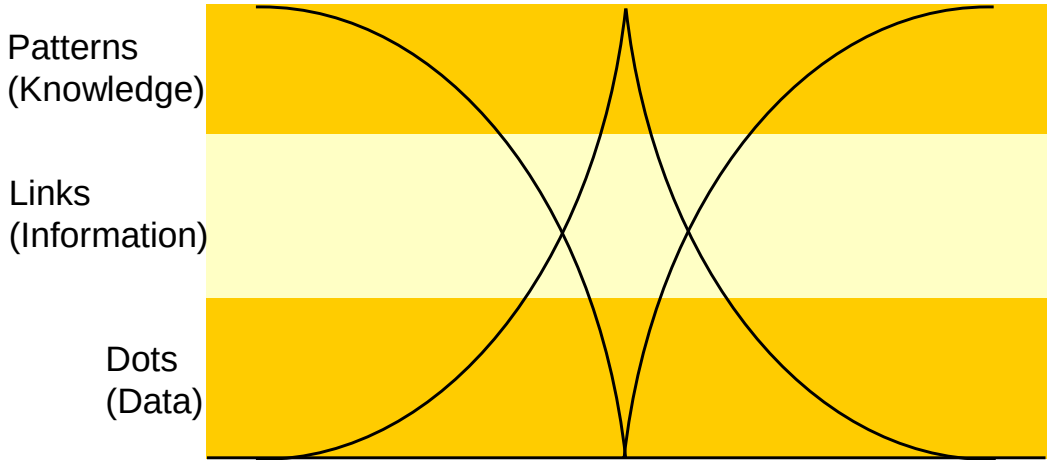


Figure 7c: Data & Pattern Processing Combined