TR-H-300

MemeStorms: A Computational Model of Human Working Memory.

Andrzej BULLER (ATR-HIP/Tech. Univ. Gdansk)

2000.6.12

ATR人間情報通信研究所

〒619-0288　京都府相楽郡精華町光台2-2-2　TEL: 0774-95-1011

ATR Human Information Processing Research Laboratories

2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Telephone: +81-774-95-1011

Fax : +81-774-95-1008

Andrzej Buller

ATR Human Information Processing Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
&
Technical University of Gdańsk
Faculty of Electronics, Telecommunications & Informatics
Ul. G.Narutowicza 11/12, 80-952 Gdańsk, Poland
buller@pg.gda.pl

March 31, 2000

MemeStorms: A computational model of human working memory

Abstract: This report presents a computational model of human working memory in which populations of contradictory memes fight for domination. As a test-bed for the model's psychological plausibility the Mouse Effect was taken. Human subjects, asked to express their feelings about a person or a social situation by the location of cursor (the center of the screen - highly positive, the border of the screen - highly negative). The subjects' feelings sometimes oscillated between a highly positive value and a highly negative value even in absence of new data about the evaluated object. The proposed model, called MemeStorms, demonstrates the same kind of oscillations. A modular structure implementable on ATR's CBM (Cellular [Automata-based] Brain Machine) of MemeStorms-based working memory has been designed.

When I was a boy of 14, my father was so ignorant
I could hardly stand to have the old man around.
But when I got to be 21, I was astonished at
how much the old man had learned in 7 years.

—MARK TWAIN

1. Processing of social concepts

As Stephen J. Read and Lynn C. Miller, University of Southern California psychologists, noted, people are constantly trying—consciously or not—to make sense of events in social interaction and such inferences are incredibly complex. They argue that a part of this complexity is reflected in the characteristics of social events that often provide multiple cues involving multiple modalities that are given simultaneously and changing over time[1]. Since it is rather difficult to analyze this aspect of human mental activity remaining in the framework of cognitive psychology, a new discipline called social cognition had to emerge. This chapter presents the principles of social cognition with its view of concept, the most representative model of social perception, as well as an extraordinary view of the roots of human social behavior.

Social cognition

Social cognition is a relatively new discipline built on a long tradition of research and theory in social psychology, as well as on new ideas and methods emerging from cognitive psychology. Hence, the basic topics in social cognition are traditionally penetrated by social psychology group stereotypes, knowledge of other individuals, and the self. Social cognition tries to penetrate the topics much deeper—to the level of basic mental phenomena. According to Daniel Gilbert, Harvard University psychologist, the discipline, when defined broadly, refers to those aspects of mental life that enable and are shaped by social experience. When

defined narrowly, social cognition refers to an intellectual movement that borrowed the techniques, theories and, and metaphors of "post-revolutionary cognitive psychology" and brought them to bear on traditional social psychological problems, such as attitude structure and change, casual attribution, social judgment, categorization and stereotyping, self-knowledge, self-deception, and the like[2]. Ziva Kunda, University of Waterloo psychologist uses the term social cognition broadly to refer to cognition, motivation and affect[3].

The social cognition movement was characterized by (a) its trust in the computer metaphor which suggested that mental phenomena are properly explained by describing a sequence of hypothetical operations and structures that might produce them; (b) its emphasis on mental representation with an attendant lack of emphasis on motivation, emotion, behavior, and social interaction; (c) its conviction that social cognition was a special case of cognition, and that theories of the former should thus be grounded in theories of the latter; and (d) its inclination for highly controlled experimental methods that maximized internal validity rather than ecological validity. While social psychology incorporated the computer metaphor, it could not accept the claim that the mind can be an all-purpose information processing device that understand social situations in the same way that it recognizes non-sense syllables or other simple objects subjects are exposed to in cognitive psychologists' labs. More convincing was the idea of modular mind having highly specialized modules, dedicated, among others, to social tasks. It is believed that a set of such "social modules" evolved as a natural consequence of the evolutionary adaptation of an organism whose survival is largely dependent on its social relations which are almost never emotion-free.

Indeed, conservative social psychology always served as the voice of conscience preventing social cognition from being marginalized[4]. As until the late 1980s most theory and research in social cognition focused on relatively "cold" cognitions involved in representing social concepts and drawing inferences from them, more recently there has been renewed interest in the relatively "hot" cognitions underlying motivation and affect which influence the way people remember and make sense of social events. This research has led to integrate cognition, motivation and affect[5].

---

[1] Read & Miller (1998)
[2] Gilbert (1999).
[3] Kunda (1999:3).
[4] Gilbert (1999).
[5] Kunda (1999:3).

Social-Cognitive View of Concept

A number of research has been done in order to learn whether the results concerning cognitive concepts apply also to social concepts. Some support to the theory-based view was provided by social psychological research investigating reliance on casual reasoning when combining concepts[6]. As for concept hierarchies and the notion of the *basic level*, confirmed experimentally within cognitive psychology, some supporting findings have been reported in reference to social concepts[7]. Nevertheless, the level that may be considered basic in hierarchies of social concepts seems far more flexible and to vary from one context to another, and more likely to be dependent on our goals than in the case of non-social concepts[8]. The messy relations among social concepts have led some to suggest that rather than consider social concepts organized in structured hierarchies, we should consider them as arranged in tangled webs[9]. This way of thinking directed social cognitivists' attention to *connectionism*—an extraordinary approach to understanding cognition[10] that started flourishing since the monumental work by David Rumelhart (psychologist, Stanford University) and his colleagues[11] has been published.

Connectionist models view knowledge representations as networks of interconnected nodes and assume that activation spreads along these connections. Activated nodes can not only activate their neighbors but also deactivate them. The positive and negative links act as constraints on the spread of activation. A positive constraint between two nodes means that the two should both be activated or deactivated. A negative constraint means that if one node is activated, the second one must be deactivated. Such models, called parallel constraint-satisfaction models, aim to satisfy as many of the constraints as possible while giving preference to the more important ones[12].

---

[6] Asch & Zukier (1984) asked people to describe individuals characterized by two conflicting traits, for example, someone who is gloomy and cheerful or someone who is strict and kind. In some cases, one trait was viewed as means for obtaining the other (a person must be strict and kind because one must be strict to protect a child), which means that a conflict was resolved through causal reasoning.

[7] For example, when describing others' personalities, people often prefer intermediate-level traits such as *kind*, which represent the highest level of abstraction still describing behavior, over most specific *charitable* or more abstract *good* which does not imply concrete behavior (John, Hampson & Goldberg 1991) (quoted from Kunda 1999:44).

[8] Cantor & Kihlstrom (1987).

[9] Andersen & Klatzky (1987); Cantor & Kihlstrom (1987).

[10] Connectionism—term introduced by Jerome Feldman, computer scientist, University of California, Berkeley, for a style of computation that emphasizes the pattern of connections in a networks of neuron-like elements (Churchland & Sejnowski 1992:461)

[11] Rumelhart, McClelland & PDP Research Group (1986).

[12] Thagard & Kunda (1998).

There are two broad classes of parallel-constraint-satisfaction models. Models using *local* representations view nodes as representing identifiable concepts or propositions—traits, stereotypes, or behaviors. Models using *distributed* representations view models as representing more basic, low-level elements so the meaning of higher-level concepts is distributed over many such nodes. So far, of the small number of connectionist models applied to social cognition, most have used local, intuitively meaningful representations[13]. It has been also suggested that distributed models may be usefully applied to social psychological issues[14].

Conectionist models also differ in how they conceptualize links and constraints. The simplest models involve only positive and negative links and constraints[15]. Such ones have been successfully applied to understanding how information from many, often conflicting sources (traits, behaviors, stereotypes) may be integrated into a person's coherent impression. Other models involve more elaborate sets of constraints (e.g., logical contradiction and simplicity)[16]. These have been applied to understand higher-level reasoning such as a jury decision making[17].

Read-Miller Connectionist Model of Social Perception

Stephen J. Read and Lynn C. Miller, University of Southern California psychologists, proposed a well-elaborated model of social cognition for which they employed a four-level recurrent neural network. The macrostructure of the model consists of (1) Feature/Input Layer, (2) Identification Layer, (3) Scenario Layer, and (4) Conceptual/Meaning Layer.

It the Feature/Input Layer incoming perceptual information about the features of social actors, features of objects, features of behaviors, and features of spoken or written language is represented as a very long vector of activations. It is assumed that hard-wired "feature detectors" capture information about curves, lines, and movement, color size, orientation and direction of movement; elementary sounds and changes in their frequencies, and then, the captured features are assembled into a set of superordinate features as black skin, wrinkled skin, white hair, unsteady gait, curved mouth, and so on. Some of the features

---

[13] Kunda & Thagard (1996); Read & Miller (1993); Read & Marcus-Newhall (1993); Shultz & Lepper (1996); Spellman & Holyoak (1992).
[14] Smith (1996).
[15] Kunda & Thagard (1996).
[16] Read & Marcus-Newhall (1993); Thagard (1989).
[17] Kunda (1999:51).

may covary with other features, as for example wide nose, black skin and kinky hair. Hence in the model one feature may automatically either activate or block another. And so, black skin may activate the part of the network that applies to kinky hair and, at the same time, block the part that apples to blonde hair.

In the Identification Layer the set of the features recognized in the previous layer is processed towards a coherent image of a person (e.g. old African American man; an attractive White woman) or an object (e.g. car) or a behavior (e.g. crossing the street). There are excitatory links between certain outputs of the Feature/Input Layer and circuits producing concepts the Identification Layer is to identify. However, there are also inhibitory links between certain circuits of the Identification Layer (which means that, for example, an activation of a circuit applying to the notion of adult female blocks automatically an activation of the circuit producing a representation of the notion 'male'), as well as links from the Identification Layer to certain circuits of the Feature/Input Layer (which means that, for example, having misidentified an individual as a woman because of "her" long hair, we may be prone to misperceive some of "her" other features).

In the Scenario Layer particular items are assembled into a scenario. The scenario pattern is defined here as a script-like representation of *who* (or *what*) *did what* (*how*, *with what*) *to whom* (or *what*), *and with what effect*. The Scenario level may be divided onto a number of plot units, where each unit concerns a basic element of a more sophisticated story. The character of the story may influence the Identification Layer in such a way that, for example, if we see someone fall, and another's hands engaging in an outward thrust, we may connect the one person's behavior to the fall of the another.

The Conceptual/Meaning Level is the place in which, based on the scenarios produced by the Scenario Layer, the inferences are made about meaning of the behavior (e.g. cooperative, aggressive, etc.), the actors' intentions and goals, characteristics of the actors, such as their traits, or the meaning of the situation. The judgments to be supposedly produced by the layer were compared with the judgments resulting from Trope's model of dispositional inference[18] and proved to be similar[19].

Read and Miller (1988) suggest that their model can be based on neural units with sigmoid-shaped activation function. They also provide a survey of the most popular methods of neural networks' learning. Unfortunately, they show no detailed description of a

[18] Trope (1986); Trope & Liberman (1993).
[19] Read & Miller (1998).

mechanism of new knowledge acquisition in their model. Nevertheless, as a general view, the Read-Miller connectionist model must be treated as a big step towards a joint biologically-psychologically plausible modeling of mind. The fundamental principles of the model are compatible with many of Gestalt principles that formed modern social psychology. Read and Miller's concept is a continuation of the way followed by S. E. Asch who argued that the processing of social stimuli was holistic[20], F. Heider who relied heavily on Gestalt principles of structure and organization[21], L. Festinger whose theory of cognitive dissonance was strongly grounded in Gestaltists' ideas of structural dynamics[22], as well as by Kurt Lewin (1935) who described person-situation interaction in terms of interacting force fields[23]. Read and Miller, adopting connectionist approach to the ideas mentioned above, seemingly made a giant step in a proper direction, however several details of their model should be elaborated towards its implementable version.

Genes vs. Memes

Neither psychological search for correlations between social behaviors nor hand-designed models demonstrating human-like social behaviors can provide an insight to the roots of the behaviors. In this matter we have no choice as to select the most convincing theory. A good candidate for such a theory is evolutionary way of thinking not necessarily limited to the domain of molecular biochemistry.

Several social behaviors are explainable in terms of gene survival. This approach denies human being's subjectivity in this play. It is argued, that the true subject is a set of genes that, although incapable of any conscious action, owing to the existence of Nature forces, can survive over generations. It sounds quite convincing that, for example, sex differences in mating behavior are not to facilitate individuals' will to have an offspring, but rather resulted from a competition among a number of alternative sorts of gene sets. Simply, when an organism, built according to a particular genetic code, demonstrated inefficient sexual behavior, the probability of its offspring appearance was relatively lower. Hence, the gene sets resulting in any inefficient behaviors usually had to die together with their "carrier". The simplest answer to the question 'Where such and such sexual and other social behaviors came from?' is that in every generation a number of sorts of gene sets were produced, and a

[20] Asch (1946).
[21] Heider (1958).
[22] Festinger (1957).

number of possible behaviors were demonstrated by organisms built according their genetic codes, but only a part of the organisms have managed to live until their reproductive age and were able to produce new gene carriers; as a result one sort of genes still exist, while other became extinct, which is equal to the fact that one sort of social behaviors still exists, while other became extinct together with the genes responsible for the behaviors.

But, the higher organism the bigger number of non-genetic factors influencing subject's social behavior (including a way mating). For instance, a rape may cause dramatic changes of victim's personality such that permanently prevent her from initiation a proper relationship with any male. This is hardly explainable in terms of gene expression, since such a change in personality acts against a given gene set's interest. Moreover, a sudden change in attitude to opposite sex is hardly explainable in terms of long-term biochemical properties of synapses. Seemingly there is a mechanism of both rapid and permanent changes in cognitive system. A quite convincing model of such a mechanism employs the notion of *meme*.

Meme—a term coined in 1982 by Richard Dawkins[24]—has been defined in a number of ways. According to Henry Plotkin, the meme is the unit of cultural heredity analogous to gene[25]. More precise is the definition by Richard Brodie who sees a meme as a unit of information in a mind whose existence influences events such that more copies of itself get created in other minds[26]. In other words, it is argued that there exist other kinds of entities whose expression is an individual's social behavior, however, the entities can, like genes, be treated as a true subject in the quest for survival. From the "point of view" of memes no matter that a rape victim blocks itself her chance for motherhood, as well as no matter that certain gene set will die because of lack of possibility to locate their copies in a new-born organism. For certain reasons memes carrying the popular belief that a rape victim losses much of her attractiveness replicates more efficiently than memes carrying another views. One can, therefore, see the history of mankind as a side-effect of a most essential process—an eternal war genes vs. memes. Genes act in a long-term time-scale that can be measured in decades. Memes can act in seconds.

What is the nature of a meme? Which way memes are stored in our minds and which way they replicate and change themselves? Based on the gene-meme analogy we may suppose that in order to replicate memes employ a genetic-like mechanism. If, for example,

---

[23] Lewin (1935).
[24] Calvin (1996: 18).
[25] Plotkin (1993: 251).
[26] Brodie (1996).

they were represented as strings of small pieces of information, certain specialized neural circuits could facilitate their crossover and mutation. This view will stop being a speculation when a meme-based model of mind is implemented and demonstrated human-like social behaviors.

Conclusion

Social psychology discovered a big stuff of counterintuitive phenomena concerning peoples' social behavior. The behavior is often irrational, to not to say—stupid. Social cognition, adapting several tools and methods developed within cognitive psychology, let social psychology start searching for mental mechanisms of the phenomena.

All social behaviors are based on judging that is more or less conscious. Hence, social judgment seems to be the most important topic in social psychology. While traditional social psychology's major question was 'How people evaluate?', social cognition's major question is 'What mental processes lead to the act of evaluation?'.

Nevertheless, based on a set of representative publications, one can have the impression that social cognition follows some hidden assumptions that are hardly acceptable. First, in social-cognitive search for the essence of concept and concept relations, the mental entities are usually treated as labeled 'blocks' an individual consciously play in his/her mind, while it is possible that majority of the 'blocks', not labeled at all, is 'manipulated' beyond the individual's control and even observation. Second, although one has to appreciate the employment of artificial neural networks to create social-cognitive models, it may be noted, that the employed networks are not the state-of-the-art ones. Old-fashioned neural networks are provided with data and produce results. It is satisfactory in case of simple cognitive processes as, say, character recognition. In case of social judgment—a process which lasts for seconds, minutes, hours, during which a subject may change his/her mind several times—the simple networks, although sometimes their stable responses (not to be confused with continuous behaviors) are psychologically plausible, in their essence are psychologically implausible. Moreover, convenience of local intuitive representations suggested as the reason for treating network nodes as labeled concepts sounds disarming.

A notable exception is the Read-Miller model of social perception based on a distributed representation of information processed in multi-layer recurrent neural network. Although the model proved its explanatory power in reference to several cases of social perception, it provides no suggestion about a way the described neural layers can learn their functions. Moreover, the model cannot use an exact remembrance of past events to make inferences about current events.

Hence the final conclusion from the chapter: In order to answer its major question, social cognition has to employ some modern tools and methods cognitive psychology—its traditional source of new techniques—has not adopted yet. Indeed, this is already happening in the framework of new emerging disciplines—Dynamical Social Psychology and Memetics.

Yet each man kills the thing he loves
By each let this be heard,
Some do it with a bitter look,
Some with a flattering word.
The coward does it with a kiss,
The brave man with a sword!

—OSCAR WILDE
from *The Ballad of Reading Gaol*

2. Dynamical Social Psychology

Despite its seductive repertoire of techniques that let one cope with objects demonstrating complex behaviors, dynamical system theory so far has been noted and employed by few social psychologists. It is not surprising when taking into account the large stuff of tools and methods computer science delivered, via cognitive psychology, to social cognition the last one has not digested yet. Nevertheless, a narrow stream of research aimed to answer the big question of social cognition in the course of using dynamical-system-theoretical methodology has gained much strength within the last decade [27]. It was noted that social psychology was not aware of the potential usefulness of certain types of data, such as time series measurements of a single variable. It was also noted, that the subject matter of social psychology is the social group, as well as the dyad, and the individual mind—all of them consisting of mutually independent, interacting elements describable in terms of dynamical systems[28]. The appearance in 1998 of the book *Dynamical Social Psychology*[29], written by Andrzej Nowak, University of Warsaw psychologist, and Ronald Vallacher, Florida Atlantic University psychologist, may be recognized as a birth of new discipline.

[27] Nowak, Szamrej, Latané (1990); Nowak, Lewenstein & Szamrej (1993); Vallacher & Nowak (1994a); Ostrom, Skowronski & Nowak (1994); Vallacher & Nowak (1997).
[28] Nowak, Vallacher & Lewenstein (1994).
[29] Nowak & Vallacher (1998a).

Flow of social judgment

Social psychologists have documented a number of interconnectedness of socially grounded thoughts within cognitive structures. Thoughts about a social role, a particular person, an ethnic group, or oneself are assumed to consist of specific elements such as traits, behaviors, characteristics, etc. that are coordinated in some fashion to represent one's thoughts. Among the proposals of coordinating structures we can mention such concepts as schema[30], category[31], prototype[32], network[33], script[34], and implicit theory[35]. Unfortunately, as Vallacher and Nowak note, psychologist have not been so dutiful in investigating the dynamic nature of thought with its tendency to evolve and change over various time-scales. The relative lack of explicit attention to the flow of thinking leaves one with a false impression that the cognitive structures detailed by psychologists are rather like fixed architectures in which all the elements fit together like pieces of stone in a pyramid[36].

As a matter of fact, the mind remains active, producing a rapid turnover in output despite the lack of environmental input, even during sleep[37] and under conditions of sensory deprivation[38]. *Whether thinking about oneself, an intimate friend, or a perfect stranger, our thoughts seem to have a trajectory of their own, changing from one set of elements to another on a rapid time scale, even when no new information is provided and there are no external pressures for updating our thoughts*[39]. This puts the topic close the ideas covered by the term *stream of consciousness* coined by William James over a century ago[40].

The new quality giving birth to a new discipline emerging from social cognition is paying homage to the Jamesian metaphor of the stream of consciousness by providing a conceptual and operational scheme for exploring the intrinsic dynamics of social judgment. By intrinsic dynamics Vallacher and Nowak mean internally generated patterns of temporal variation that occur in the absence of external forces. They began by developing a rationale

---

[30] e.g. Rumelhart (1980); Taylor & Crooker (1981).
[31] e.g. Rosch (1978)
[32] e.g. Cantor & Mischel (1979); Posner & Keele (1968).
[33] e.g. Anderson & Bower (1973).
[34] e.g. Schank & Abelson (1977).
[35] e.g. Wegner & Vallacher (1977).
[36] Vallacher & Nowak (1994b).
[37] Hobson (1988).
[38] Zubek (1969).
[39] Vallacher & Nowak (1994b).
[40] James (1890/1950).

for this exploration, suggesting why social judgment is inherently dynamic and why it may be important to consider it in these terms[41].

The key concepts of the Complex Dynamical System Theory are (1) set of state variables, and (2) order parameter. The set of states, depending on the level of consideration, can be represented by either activations of neurons constituting our nervous system, or a bundle of separate and unrelated images, event memories and trait ascriptions. Since investigating histories of changes of particular state variables is impractical, the idea of order parameter has been invented. What in thinking about someone can play the order parameter's role? Vallacher & Nowak suggest that our thoughts tend to have a "bottom line" to them, some sort of provisional integration that reflects our general sense of the person, and that evaluation is an obvious candidate for this role[42]. If even the specific issue is impression formation, political attitudes, or assessments of responsibility, in each of the cases the broader issue can be framed in terms of how people feel about the person or topic at stake. Even if, sometimes, more specific dimensions of judgment are of interesy (e.g., fairness, intelligence, social skill), it is hard to identify dimensions that are devoid of evaluation[43].

One of possible ways of evaluative integration of diverse cognitive elements is based on the assumption that each element (i.e., thoughts, attributes, and emotions concerning the target of judgment) is valenced and that a summary judgment—an evaluation—at each iteration represents some computed function (e.g., weighted average) of the valences associated with the activated elements[44]. It follows that as the configuration elements changes over time, there is a concomitant potential for change in one's overall evaluation of the target of judgment. The variation over time in some index of the evaluation is called by Vallacher and Nowak the *stream of social judgment*.[45]

The Mouse Paradigm

It is not a simple task to gain access to the stream of social judgment. How to get people to express their feelings continuously, but do so without reporting on them? To solve

[41] Vallacher & Nowak (1994b).
[42] Anderson (1981); Fiske & Taylor (1991); Wegner & Vallacher (1977) quoted by Vallacher & Nowak (1994b).
[43] Kim & Rosenberg (1980).
[44] Anderson (1981).
[45] Vallacher & Nowak (1994b);

the paradox, the researchers took a classic idea of social psychology, as well as welcomed the aid of modern computing technology[46].

Almost fifty years ago it was suggested that evaluation could be considered an implicit approach-avoid response. In other words, a judge's preferred proximity to a target represents an expression of his or her current feeling about a target. The closer judge's preferred distance from the target, the more positive his/her feeling[47]. As Vallacher and Nowak note, a movement toward or away from the target may represent change in judge's feelings about the target. In order to capture the movements, they invented a computational technique called *Mouse Paradigm* and used it to investigate the streams of people's social judgments[48].

The Mouse Paradigm takes its name from the "mouse" connected to a personal computer as a part of user's interface. Below the experimental procedure described by Valacher and Nowak is provided:

> ...On the computer screen two objects are presented: an arrow reflecting the position of the cursor and a small circle positioned in the middle of the screen. The arrow is said to represent the subject, and the circle is said to represent a particular target of judgment. Subjects read a description of a target person or of an event involving themselves and a target person and are asked to think about the target. As they do so, they adjust the arrow in relation to the target circle (by moving the mouse) so as to express their moment-to-moment feelings about the target over a 2-min period. The mouse is positioned on the side of the keyboard corresponding to subject's dominant hand.
>
> In introducing the task, the experimenter informs subjects that if they feel positive about the target, they should move the arrow toward the circle by moving the mouse; by the some token, if they feel negative toward the target, they should move the arrow away from the target. The experimenterthen informs subjects that if their feelings about the target change, they should move the arrow toward or away from the target to express these changes. Subjects are free to adjust their position relative to the target as often as much as is necessary to reflect their feelings about the target as they continue to think about him or her.
>
> After a 20-s practice session in which subjects moved the mouse and observed the corresponding movement on the screen, the screen cleared and a description of a

[46] Vallacher & Nowak (1994b).
[47] Hovland, Janis & Kelly (1953) quoted by Vallacher & Nowak (1994b).

particular target person appeared. Subjects then began the 2-min mouse procedure. The location of of the arrow was assessed 10 times per second for a total of 1,200 potential data points. Research to date reveals that all subjects spend the first few seconds moving from initial position (immediately adjacent to the target) to a "safe" starting point. For this reason, subject's movements during the first 3 s were not included in subsequent analyses. The program preserves the Cartesian coordinates of each data point, although for purposes of our initial investigations (described below), only the absolute distance from the target was considered. This distance provides a measure of subjects' moment-to-moment feelings about the target.[49]

It has been experimentally confirmed that social judgment shows temporal variation in the absence of new information or social influence. As for temporal patterns of people's feelings, one can argue, that the mind is so busy place the turnover in thought that people experience result from a noise obscuring a more stable signal. The Mouse Paradigm was used to establish the counterintuitive hypothesis that temporal variation in judgment could be meaningful.

Astonishing Oscillations

The aim of initial Vallacher & Nowak's study was to uncover temporal patterns in social judgment[50]. Below the description of their experiment is provided:

> ...Nine subjects performed the mouse task for each of four hypothetical event descriptions (presented in random order). Each description was designed to engender some ambivalence in subjects so to maximize the likehood that their moment-to-moment feelings would show temporal variation (i.e., fluctuation between positive and negative feelings). The descriptions can be summarized as follows.
>
> 1. The subject meets an attractive person of the opposite sex at a party and arranges to date the person the following weekend. Later, the subject overhears another person tell someone else that the subject's future date had once dated someone who had tested positive HIV but broken off the relationship once he or she had discovered this.

---

[48] Vallacher & Nowak (1992).
[49] Vallacher & Nowak (1994b).
[50] Vallacher & Nowak (1992).

2. The subject is discussing marriage plans with his or her prospective spouse. The marriage partner discloses that there is a history of a potentially fatal genetic disease in his or her family and that the odds are 1 in 4 that any offspring they have will develop the disease.

3. The subject is having an increasingly heated argument with his or her marriage partner concerning their relative contributions to the household. One of them (unspecified) storms out the room.

4. The subject learns that a close friend once stole money from another of his or her friends. The close friend never admitted to the theft or tried to make amends to the victim.

In thinking about each description, subjects were instructed to indicate their moment-to-moment feelings about the event and/or terget (i.e., the date, the impending marriage, the spouse, the close friend).

Results revealed intrinsic dynamics in all cases, with judgment corresponding to one of several distinct temporal patterns.[51]

Figure 2.1A-D presents the raw data generatet by four subjects. The horizontal axis represents time (0-120 s), the vertical axis the absolute distance from the target in pixels. Visual inspection of the plot in the Figure 2.1A suggests that the subject's judgments alternated between two values. Autocorrelation analysis revealed the plot's periodic nature. It's Fourier analysis revealed one dominant low frequency. In contrast, the plot in Figure 2.1B has a highly positive value for the first minute, and then, suddenly collapses to near one third of its maximum value and remains around this level. Autocorrelation analysis of this plot suggests that a judgment in one point in time can be a good predictor of a long series of judgments. The Fourier transforms of the subject's data reveals only a dominant zero frequency, confirming the lack of periodicity. The plot in Figure 2.1C shows seemingly irregular oscillations with decreasing amplitude (as if judgment were converging on a stable attractor) and suggests no predictability. Autocorrelation and Fourier analysis of the plot in Figure 2.1D suggest its chaotic nature.[52]

[51] Vallacher & Nowak (1994b).
[52] Vallacher & Nowak (1994b).

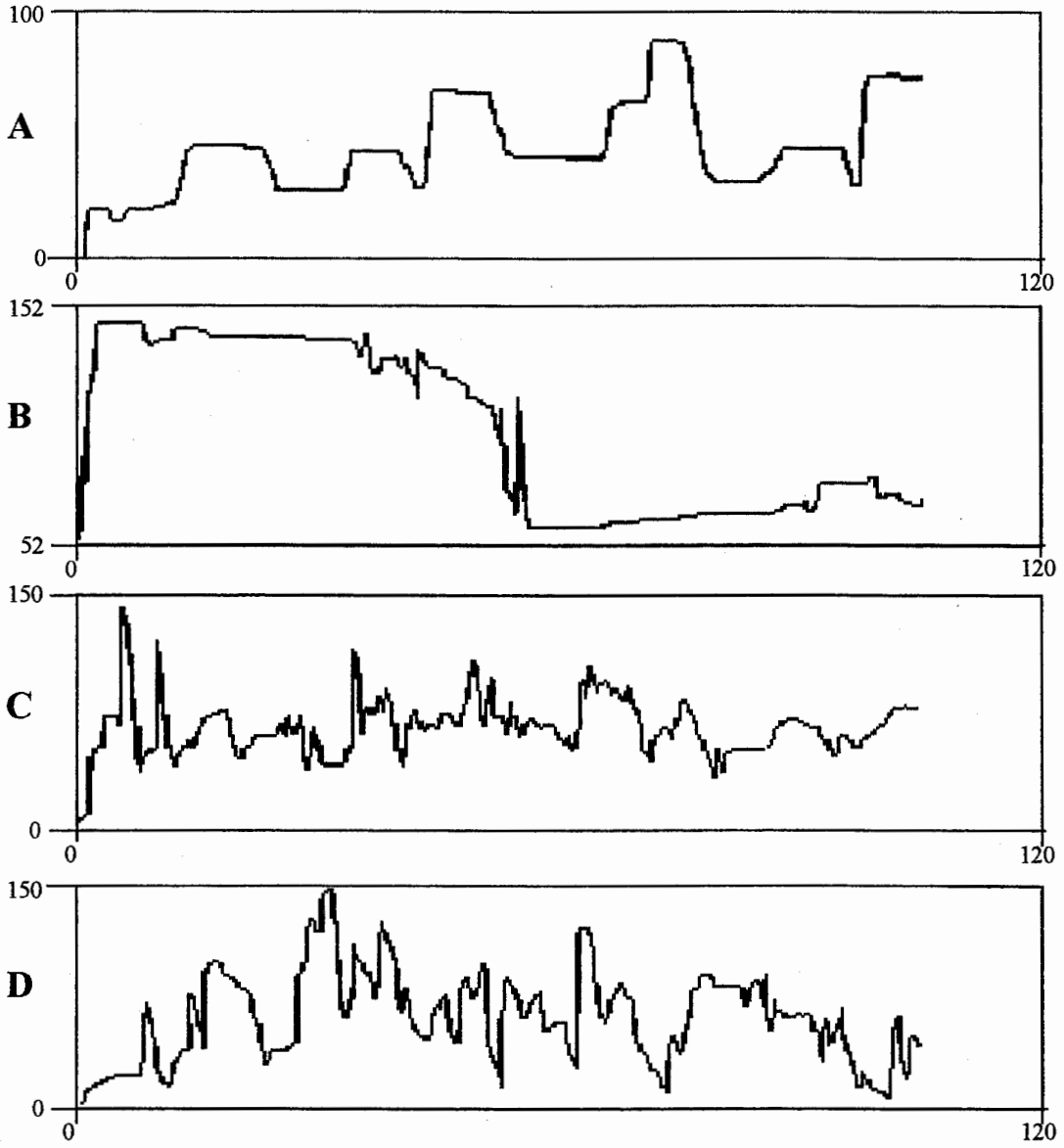

Fig. 2.1A-D. The Mouse Paradigm. The data have been generated by four subjects who used a mouse to express their feelings about a target . The horizontal axis represents time (0-120 s), the vertical axis the absolute distance from the target in pixels (Adapted from Vallacher & Nowak (1994b)).

For Vallacher and Nowak the above results strongly indicate that variation in judgment over time is not simply a noise obscuring "true" stable values, but rather represent the essence of the judgment process itself. Thoughts unfolding in accordance with temporal patterns strongly suggest that this mental activity is a result from a process in an underlying dynamical system.[53]

[53] Nowak & Vallacher (1998: 99).

Conclusions

Social judgment is seemingly the most important of mental activities of contemporary human. The quality of our evaluation of encountered persons or social situation determines our career and, often, our chances for survival. Hence, an extremely challenging topic in modern psychology is to investigate the mechanisms of social concept processing that lead to evaluation of persons or social situations. Neither traditional cognitive psychology nor traditional social psychology could provide answer for the question 'How social concepts are processed in the human mind?' Cognitive psychology – because it concentrates on investigating only basic mechanisms of perception, memory and attention, social psychology – because it used a poor methodological toolkit and falsely assumed that human social behaviors could be predictable based on statistical analysis.

A theoretical and methodological breakthrough was DSP (Dynamic Social Psychology) that considers human social behaviors and their underlying mechanisms in terms of changing state of mind understood as a complex dynamic system. A computational technique invented in the framework of DSP by Robin Vallacher and Andrzej Nowak, called *Mouse Paradigm*, allows us to observe a plot of subject's changing feelings about described persons or situations. Owing to a series of experiments using the Mouse Paradigm, there is empirical evidence that people's feelings about an object or a social situation can oscillate, switching sometimes from strongly positive to strongly negative and back. DSP took from Complex System Science several theoretical constructs, as state space and attractor, facilitating analysis of dynamics of people's feelings. A new revised meaning of social behavior predictability emerges from the dynamics. However, there is still no model of human mind, which would propose a vision of detailed mechanisms causing social judgment varying in time.

An attractive candidate for a basis for such a model seems to be a five-element memory system suggested by Endel Tulving. The five elements seemingly appear consecutively during filo- and ontogenesis. They are Procedural Memory, Perceptual Representation System (Filtering Memory), Semantic Memory, Working Memory, and Episodic Memory. Based of Tulving's proposal I argue that Working Memory plays an integrating role in human mind as a coordinator of all other memories' activities. I also argue, that dynamics of social judgment results immediately from dynamic properties of Working Memory itself. Hence, the second part of the dissertation will be devoted first of all to a model of Working Memory.

There are several models of Working Memory based on different key concepts. Among computational models we can distinguish (1) a group describing mental phenomena as generating particular actions in the course of using symbolic rules, as well as (2) a group considering mental process as parallel exchange of generally meaningless signals among a number of distributed modules of mind. In the second case a kind of computation in the brain is assumed, however, it has nothing in common with a naïve describing of mind as a kind of traditional computer with a memory unit and a separate processor. I argue that dynamics of social judgment can be explained using a new kind of model of Working Memory that combines parallel distributed processing with symbolic rule-based paradigm. More precisely speaking, I propose a model in which a rule is a instance of meme—a unit of information capable of replicating itself, while a connectionist network, to be in major part evolved, facilitates meme interactions.

Contemporary computer engineering offers a number of excellent tools facilitating cognitive modeling, namely, artificial neural networks, genetic algorithms and cellular automata—all of them being instances of a more general class of self-organizing systems that are a subclass of Dynamic Complex Systems. An attractive tool that can be used for an implementation of a large-scale neural model of a thinking brain is CBM (Cellular-Brain Machine). Using the CBM one can simulate evolution of neural networks that grow in cellular automata workspace. Higher-level brain architectures, including a neuro-computational model of Working Memory, can be build of the evolved modules.

The Way is shown as five books
concerning different aspects.
These are Ground, Water, Fire,
Tradition (Wind), and Void[54].

—MIYAMOTO MUSASHI
from *A Book of Five Rings*

3. Mandala of Mind

In this chapter I formulate a proposed model of mind as a sort of initial guidelines for artificial brain building. The artificial brain is to demonstrate a great deal of mental skills, including object recognition, symbolic reasoning, action planning, language acquisition, and, as the most important, social judgments. The artificial brain is to show that a mind, just the same a humans have, can reside in a device human can made of non-organic stuff. Hence, although this chapter does not deal with empirically confirmed mechanisms of human mind, it is not to be treated as a pure speculation. For it is an introduction and technical description of something being built to behave as it had the same mind as human and help us in investigating the true, still hidden source of the phenomenon of thinking—the source that Nature has utilized in one of many possible ways.

Basic Assumptions

The proposed model of mind is grounded on a set of assumptions I formulated based on the most convincing and empirically proved theories, both from the field of psychology and other relevant fields. These assumptions concern: modularity, dynamics, self-organization, and freedom of implementation.

[54] Miyamoto (1645/1982: 43)

I. Modularity. A mind can be analyzed in terms of cooperating special-purpose modules. This view emerged in 80's as an alternative to the model treating the brain-mind system as unified general problem solver and is currently accepted by the most influential cognitive scientists.

II. Dynamics. A mental process resulting in the emergence of a particular subject's decision or judgment consists in time consuming changes of the state of mind, where the state is understood as an informational content of particular modules of the mind. Dynamic approach to cognition and emotion, being an alternative to the traditional view of mind where stimuli were processed towards particular reactions according to a particular algorithm, let us employ a well-developed methodology of control theory and neural engineering.

III. Self-organization. A dynamic system in certain circumstances (matter/energy supply + initial state far from equilibrium) is able to increase itself its complexity understood as a function of the number of elements constituting the system and their mutual relationships. Usually, the higher complexity of a given system, the richer repertoire of demonstrated behaviors. This is the phenomenon of self-organization that has been observed in physics (e.g. Bénard cells), in chemistry (e.g. BZ-reaction), or in biology (primitive organisms form collective bodies that demonstrate much richer repertoire of adaptive behaviors than they could do acting separately). In all observed cases of self-organization there is no target pattern a system could aim at. A current pattern is a function of its precedent pattern. Pattern-to-pattern transition is a function of properties of elements constituting the system. Although the elements' properties can be defined in a simple way, the ultimate pattern can be amazingly complex and unpredictable. Hence, armed only with contemporary research tools, we can only see the ultimate pattern when all consecutive steps of the process itself or its high-quality simulation are done. The phenomenon of self-organization takes place also in the systems created in computer's memory, which is exploited in the form of artificial neural networks, genetic algorithms and cellular automata. Since self-organization is present in so many domains of reality, why it could not concern such entity as mind!

IV. Freedom of implementation. To defend this assumption is the most difficult matter. Popular view states that one of the necessary preconditions for the phenomenon of thinking is a brain consisting of nerve cells built of proteins. On the other hand, there are no other reasons for such "proteinocentricism" as the fact that nobody saw a thinking entity equipped with non-proteinaceous brain. After all, 'nobody saw' does not mean that 'nobody cannot see'. Perhaps one day we will be able to demonstrate a silicon brain thinking the same way as human's brain. For the support for this statement let us consider the following argumentation:

if self-organization is omnipresent, and we have a set of M initial replicators (entities that can reproduct themselves) labeled R_j, j = 1 .. M , then each of the replicators will be a subject to a series of transfigurations towards one of N possible forms F_{ij}, i = 1 .. N, including *nonexistence*; Each of the forms is a function mapping a set of stimuli, labeled S, onto a set of resulting behaviors, labeled $B(S, F_{ij})$. Computer simulations of evolutionary processes shown that for a given S and (i,j) such that F_{ij} is a non-trivial form, an $F_{pq} \neq F_{ij}$ such that $q \neq j$ and $B(S, F_{ij}) \approx B(S, F_{pq})$ can be evolved. In other words, from two different initial replicators, through two different evolutionary paths, two forms demonstrating almost identical reaction for given stimuli can appear. What about the case when F_{ij} is a human brain, while R_q a silicon or virtual replicator? Can $B(S, F_{ij}) \approx B(S, F_{pq})$? Nobody knows. But if the possibility were rejected *a priori*, it would be hardly to call such an act a scientific procedure. On the contrary, if somebody's theorem is that something is impossible, it seems to be good idea for a scientist to try to find a refuting example. In case of the assertion that only an organic matter structured onto neural network can host a mind, one of ways of refutations is to build an artificial brain and show that it can think as a human. The first step is to formulate a sensible model.

Specific Assumptions

Even in the framework of the four Basic Assumptions a computational model of self-organizing mind can be build in several ways and, from the scientific point of view, there is no cues giving a priority to a particular way. Hence, further specific assumptions are subject to researcher's free choice. Nevertheless, for the sake of the project's success, it is reasonable to take into account some technical matters. The model of mind should be implemented on available equipment. The specific assumptions, therefore, has been a matter of invention confined by implementability. These are: Memory-Based Approach, Memetic Paradigm, and Cellular-Automatic Paradigm.

V. Memory-Based Approach. The notion of memory is understood here as a structured device storing and processing information that an individual acquires in the way of perception of his/her environment, as well as perception of his/her own mind/body. The Mind is understood as flow of continual changes of state of memory system. The changes manifest themselves in particular subject's behaviors. The assumed modularity of mind applies also to memory. I took arbitrarily, as the most convincing, the five-element taxonomy of memory suggested by

Endel Tulving[55]. Hence, as the basic modules of memory are propose Procedural Memory, Filtering Memory, Episodic Memory, Semantic Memory, and Working Memory. The last one differs significantly from all others memories, as its assumed function requires relatively high frequency of state changes. While slow changes of other memories can be explained as biochemical changes of properties of synapses in related neurons, this mechanism seems to be completely inadequate in case of Working Memory we make responsible for fast memorization of complex perceived patterns and shuffling them when planning sophisticated actions. Therefore, to describe a dynamics of Working Memory we must enter the level of immediate interactions of various pieces of information.

VI. Memetic Paradigm. Informational interactions in the brain, as complex-system theorists argue, can be equal to superposition and phase-transition of signal oscillations in neural circuits[56]. An alternative to this proposal is the idea of navigating entities, called here *micro-assertions*, I introduced in 1990[57]. Considering changes in memory in terms of micro-assertions puts the model into the realm of memetics. Since the definitions of meme quoted in the chapter 6 may confuse memes with their perceivable expressions, I proposed to define a meme as "a unit of a cerebral code representing a single signal, or a word, or a sentence, or a rule, or a plan, or a feeling, or a verbal or non-verbal idea, which can interact with other memes in a course analogous to genetic interactions, as replication, mutation and cross-over[58]. Such view of memes seems to provide a united framework for a synthesis is of the idea of mind as a society of interacting agents[59], the proposal by William Calvin of a hexagon-based neural workspace in which populations of memes grow and fight against each other for domination in the workspace[60], and the psychological view of WM with its integrative role. As for a structure of a meme I propose an interpretable constellation of spikes in the neural circuit. Let the term microassertion be reserved for memes representing either a sentence, or a rule, or a plan.

VII. Cellular-Automatic Paradigm. All information processing in the proposed model will be described in terms of cells that change their internal states based only on measured states of neighboring cells. Such approach fits definition of cellular automata. Here the cellular-automatic paradigm will concern several levels of information processing in such a way, that

---

[55] Tulving (1995).
[56] Kelso (1995), Haken (1996).
[57] Buller (1990a).
[58] Buller & Shimohara (1999); Buller & Shimohara (2000); Buller, Nowak & Shimohara (2000);
[59] Minsky (1987)
[60] Calvin (1996).

a single cell on higher level is a cellular-automatic system of lower-level cells. This way one can build a model of any object existing in Universe, however, a preciseness of the model depends on number of employed cells. If it is to be a model for dynamic process simulation, its preciseness depends of the number of cells supported by available hardware. I decided to follow cellular-automatic paradigm, since electronic technology offers recently a hardware supporting over 890 million lowest-level cells.

"4 + 1" Memory Model

Having assumed modularity of mind and memory-based approach, I took arbitrarily, as the most convincing, the five-element taxonomy of memory suggested by Endel Tulving[61]. Hence, as the basic modules of memory I propose four long-term memories and Working Memory. Hence the proposed model I christened "4 + 1". The long-term memories are: Procedural Memory, Filtering Memory, Episodic Memory and Semantic Memory. Working Memory has been distinguished, since it differs significantly from other four memories. Its primary task is to facilitate thinking understood as evaluating perceived situation and planning behaviors. Other memories play only a supporting role of high capacity storages of appropriately structured information. The diagram depicting the "4 + 1" memory model takes the shape of a mandala in order to turn readers' attention to its multiple symmetry and universal harmony (Fig. 3.1).

Each of the elements constituting the "4 + 1" model is a complex dynamical system perceiving one data, processing them in a non-trivial way, and producing another data. The data production is interpreted as a given device's behavior. The non-triviality of data processing results from the working assumption, that it is state-based. How to present the "4 + 1" memory model as State-Feedback Diagram (SFD)? How can State, Function1 and Function2 represent a performance of the five particular kinds of memories?

[61] Tulving (1995).

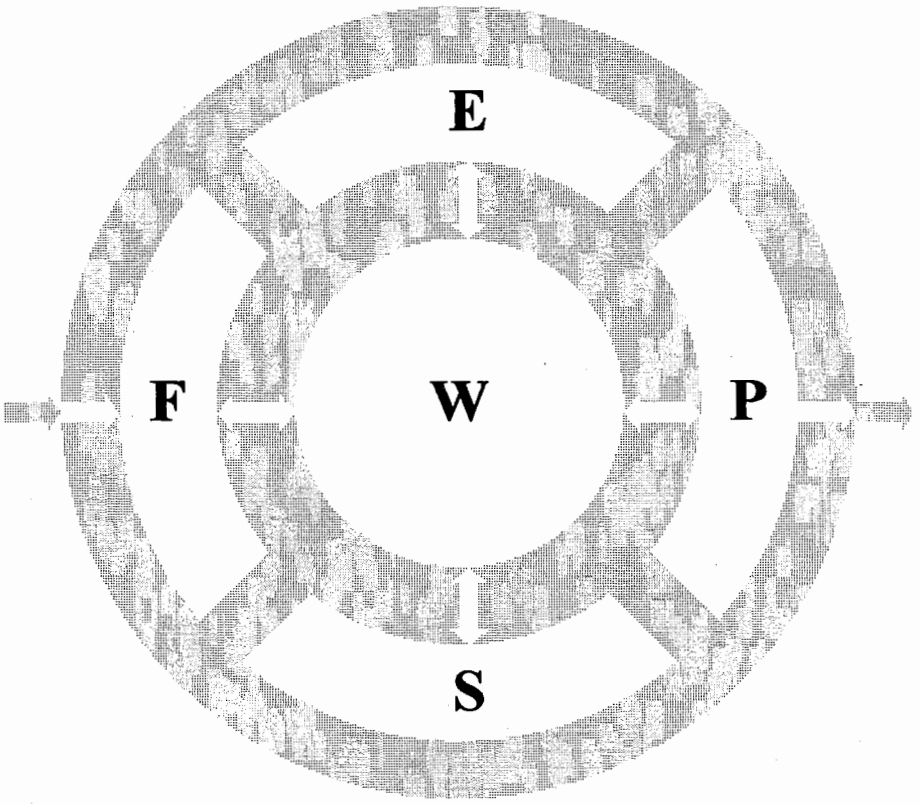

Figure 3.1. The "4 + 1" Memory Model as a Mandala of Mind.
F – Filtering Memory, P – Procedural memory, E – Episodic Memory,
S – Semantic Memory, W – Working Memory.

Figure 3.2 is a proposal of a synthesis of the state-and-function-based general diagram of any system and the memory-based view of mind. The function changing the system's state is divided onto five functions returning new states of all of the five kinds of memories. The new states are in four cases (Working, Episodic, Semantic and Procedural Memory) is calculated based on the old states of Working Memory, while in case of Filtering Memory, its new state is calculated based on currently perceived data, as well as on the old state of Working Memory.

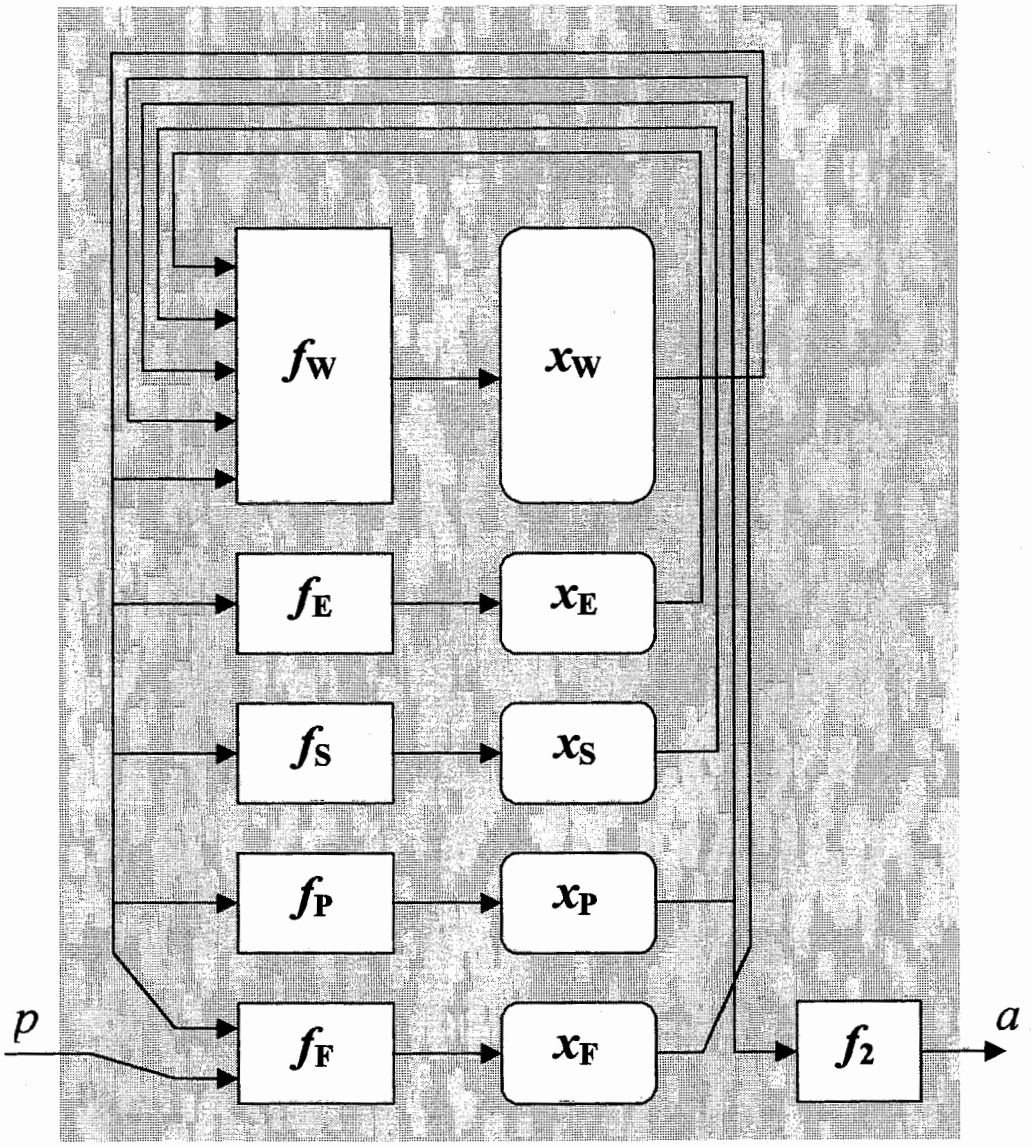

Fig. 3.2. A functional diagram of the "4 + 1" memory system. x_W, ..., x_F – states of Working Memory, Episodic Memory, Semantic Memory, Procedural Memory, and Filtering Memory, respectively; f_W – a function that for an old state of Working Memory, Episodic Memory, Semantic Memory, Procedural Memory, and Filtering Memory, respectively, return a new state of Working Memory. f_E, f_E, f_P – functions that for an old state of Working Memory return a new state of Episodic Memory, Semantic Memory, and Procedural Memory, respectively; f_F – a function that for an old state of Working Memory and perceived data p returns a new state of Filtering Memory; f_2 – a function that for a given state of Procedural Memory returns an action a.

Comparing with other memories' functions, Working Memory's function requires relatively higher frequency of state changes. While slow changes of long-term memories can be explained as biochemical changes of properties of synapses in related neurons, such mechanism seems to be completely inadequate in case of Working Memory I make responsible for fast memorization of complex perceived patterns and shuffling them when planning sophisticated actions. Hence, the idea to consider the state of Working Memory in terms of current constellations of impulses appearing and disappearing in a high frequency in neural circuits constituting this kind of memory. A constellation preserving its identity for a considerable period of time can represent a meme. Consecutive changes of Working Memory state may be equal to meme interactions leading to self-organization of higher mental structures. The same may apply to other kinds of memories, according to the assumption about the freedom of implementation.

Summary

Seven assumptions have been formulated to confine the set of possible ways of building a working model of mind. The first four, called basic assumptions, are based, among others, on the most convincing and empirically proved theories worked out in the field of both psychology and other relevant fields. They assumptions concern: modularity, dynamics, self-organization, and freedom of implementation. The next three ones, called specific assumptions, express an arbitrary decision about paradigms making the idea of mind modeling technically realistic. The assumptions concern: memory-based approach, memetic paradigm, and cellular-automatic paradigm.

A memory model called "4 + 1", consisting of Procedural Memory, Filtering Memory, Semantic Memory and Episodic Memory—each of them coordinated by Working Memory, has been selected for further consideration and transformed to a form compatible with a general scheme of dynamic system, i.e. described in terms of states and functions. Owing to this, we have a technical framework for an implementation of each of the memories as a platform for meme interactions leading to self-organization of higher mental structures. In case of Working Memory the memes can be represented by certain constellations of impulses appearing and disappearing with high frequency in neural circuits constituting this kind of memory.

Navigare necesse est.
Vivere non est necesse.

—PLUTARCH

4. Society of Memes

In the previous chapter, in the framework of the *VI*-th specific assumption (freedom of implementation), I adopted the term *meme* and introduced its specific instance called *microassertion* that can contribute to a sentence, or a rule, or a plan. In this chapter I will propose a meme structure, as well as principles of meme behavior in Working Memory.

Atoms of thought

An act of perception of any object or situation finishes in recognition and evaluation of the object or situation. Both recognition and evaluation means that a given percept has been associated with a number of notions or with an immediate action. Let us consider two examples perceptually grounded behaviors (Table 4.1 and Table 4.2).

A subject sees a flower. The percept is momentarily associated it with such notions as "tulip", "fresh", "not too big", "something my sister likes", "strange color", and many others at the some time. An immediate effect may be that the subject examines spontaneously the flower's smell and sets mind on buying some. However, although a membership of the perceived object to the set of tulips is beyond discussion, it may be not obvious that the flower is fresh. The subject, when asked whether the tulip is really fresh, may postpone his/her answer and, pushed, may say reluctantly "I suppose... hmm... rather fresh". Moreover, the same subject, after buying some tulips because they seemed fresh, may regret to buy old tulips. The change of view can take place without new data in the matter.

Table 4.1. First example story

A subject is determined to follow the rule: "I can have a date only with somebody who is nice and rich". Once he/she meets somebody who proposes a date. Visible clues provide evidence that the proponent's richness is about 60% of assumed standards, while, according to learned criteria, only 40% of the features taken into account let the proponent to be labeled "nice". The subject hesitates. "To agree or not to agree?"

Table 4.2. Second example story

Why in some cases people evaluate decidedly, while in other cases reluctantly? Why their views may change even in absence of new data? What mental mechanism is responsible for this very humanly feature of mind? Nobody knows and still cannot know. But this does not mean that there is no way how to build an artificial entity demonstrating reluctance. The simplest way is to employ classic fuzzy set theory with its well-formulated logic. The problem is that set-theoretic formulas impose symbolic knowledge representation and traditional-computer-style calculation. It also must be noted that classic fuzzy calculus is an act, while we need a process. Let us therefore consider an instance of the *Society of Mind*, described briefly in the Table 4.3.

1. subject's feeling is represented in Working Memory as neither a single entity nor a single event, but as a coincidence of a number of entities or events appearing at the same time in several places,
2. some entities/events can concern a particular feeling, while other, appearing at the some time in the subject's Working Memory, can concern an opposite feeling, and
3. the contradictory representations can fight for domination over the Working Memory's space and subject's current feelings/beliefs depend on which of the populations has been successful for a certain period of time. This original idea provides a non-trivial solution for the problem of representation of fuzziness, as well as grounds for a non-trivial model of dynamics of subject's feelings.

Table 4.3. An instance of the *Society of Mind*

What is the entity or event that alone means nothing while as a member of successful population in the Working Memory it contributes to subject's feeling? Searching for the frames of mind I coined the term *microassertion*. When a population of microassertions

referring to the proposition "This tulip is fresh" becomes a dominating population in Working Memory, the proposition "This tulip is fresh" becomes subject's final belief—a result of a process of perception/categorization that took some time and could follow a bizarre course. The same applies to subject's decision about a purchase of a book, or to a decision to say "Why not?" to date proposal. Having employed the notion of microassertion as an atom of thought, and considering mental phenomena in terms of behaviors of populations of microassertions, the "4 + 1" model copes with the situation when cues contributing to subject's judgment are not clear or contradictory.

Streams of perception

Let us return to the subject determined to have a date only with somebody who is nice and rich. In the language of logic the subject's principle could take the form "**I can have a date with X, if X is nice and X is rich**", however such an aggregate description applies to the case when for the subject richness and beauty are of the same significance. How to represent the principle in such case when, say, beauty is important but not as much as richness? Norman Anderson provided a formula in which the evaluation was the weighted average of all the elements in the structure, with the weight of each element corresponding to its importance[62]. Unfortunately, Anderson's formula can provide only a constant value of an evaluation. The solution I propose can provide the solution varying in time even in case of constant input data. It uses logic, but pretty far from classic logic.

Let as assume that the subject is just two times stronger impressed by richness that by beauty, which may be declared by the subject him- or herself or inferred from his/her behavior. When a date is proposed, a stream of memes contributing to the subject's awareness that a date is proposed enters Working Memory, from there get in touch with Semantic Memory, and, as a result of this action, the Semantic Memory starts releasing several kinds of memes related to the topic "date". Among the released memes there are copies of the microassertion of the first kind containing the proposition "**Agree** [for the date with X if X is] **nice**", as well copies of the microassertion of the second kind containing the proposition "**Agree** [for the date with X if X is] **rich**". The number of microassertions of the first kind released for a certain period of time may be a half of the number of microassertions of the second kind released at the same time, which reflects the relatively smaller importance of richness. Let us note that in a classic logical system this way the differentiation of

importance between two conditions in a rule would make no sense. A formula "Y if X_1 and X_2" cannot be replaced with the pair {Y if X_1, Y if X_2} because the pair is an equivalent of "Y if X_1 or X_2". But the "4 + 1" model deals with populations of microassertions, hence, the classic meaning of *and* and *or* becomes inadequate, at least in reference to such sophisticated mental phenomena as social judgment. Such a microassertion as, say, "**Agree** [for the date with X if X is] **nice**" is not to be used for an immediate production of a subject's decision, but to contribute, together with a number of sister-microassertions, to the subject's decision.

As it has been proposed, according to criteria the subject acquired in the past, only 40% of the features taken into account let the date-proponent to be labeled "nice", while some visible clues suggest a 60% certainty that the proponent's richness is satisfactory. The subject's hesitation "To agree or not to agree?" can take place even when the perceived data constituting a basis for a judgment or decision remain unchanged. Oscillation in social judgment in absence of new data has been experimentally confirmed[63]. Nevertheless, roots of the phenomenon still remains an open question. The "4 + 1" model provides an extraordinary explanation.

The proposed solution consists in that when cues causing contradictory conclusions or feelings are perceived, a sensorium, in cooperation with a Semantic Memory, produces streams of memes representing the contradictory data and direct them to Working Memory. And so, since in the above example just 40% of the perceived features qualifies the date-proponent as "nice", a number of memes contributing to the belief that "[the date-proponent is] **nice**" and arriving to the working memory in a period of time will be 2/3 of the number of memes contributing to the belief that "[the date-proponent is] **not nice**" arriving in the same time. The stream of the memes contributing to the belief "[the date-proponent is] **rich**" will be 3/2 times more dense that the stream of memes contributing to the suspicion that "[the date-proponent is] **not rich**" (Fig. 4.1). Gradually Working Memory is getting full of contradictory memes.

Meme interactions

Memes after its arrival to Working Memory do not stop their motion. Working Memory is a physical space. The memes roam all over the space and meet other memes.

---

[62] Anderson (1981), quoted after Nowak & Vallacher (1998: 161-162).
[63] Vallacher & Nowak (1994b); Nowak & Vallacher (1998:97)

Some of the meetings result in birth of new memes. Meme *mating* is the most important, but not the only kind of meme interaction.

Let us consider Working Memory populated by the special sort of memes we call microassertions. A given microassertion can represent either a fact or a rule, however it itself is neither fact nor rule. A microassertion, together with a number of identical microassertions, can only contribute to subject's feeling that a given fact takes place or that a given rule works. Nevertheless, like assumed facts and rules, microassertions are subject to logical reasoning being manifested by a sort of mating. And so, a male microassertion "[X is] **nice**" meets the "female" microassertion representing the rule "**Agree** [for the date with X] **if** [X is] **nice**", mates "her" and, as a consequence of this union, the microassertion representing "**Agree** [for the date with X]" is born. On the other hand, "**Agree** [for the date with X] **if** [X is] **nice**, when mated by either the "[X is] **not rich**", gives birth to the microassertion representing "**Not agree** [for the date with X]".

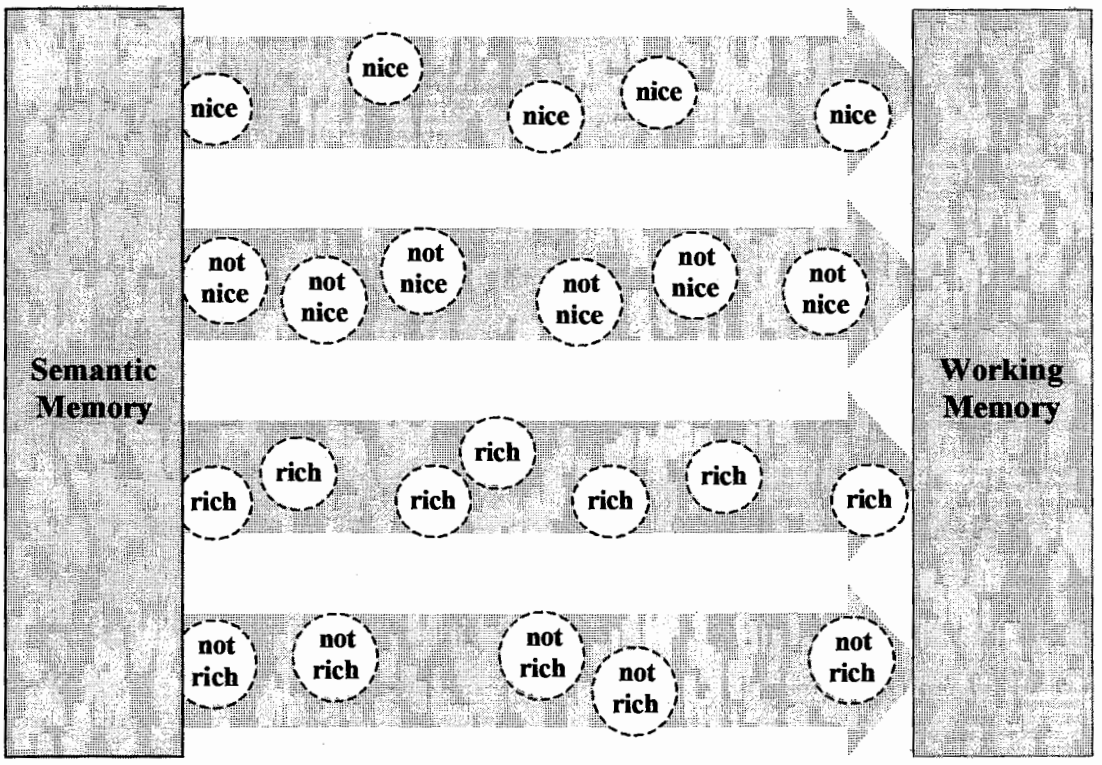

Figure 4.1. Streams of perception. In case of uncertain cues contradictory memes enter Working Memory. A value of membership of perceived object to a fuzzy cathegory is a function of the proportion between the density of a stream of memes representing a given cathegory and the density of a stream of memes representing an opposite category.

Since memes can navigate and the streams of contradictory memes are continuously supplied, the acts of mating take place at the same time in several places, which results in appearance and co-existence of populations of contradictory microassertions in Working Memory. But the co-existence in by no means peaceful. Each of the populations tries to dominate over the Working Memory. If one of them has managed to dominate for a period time, the subject's belief will such as the fact represented by members of victorious population. And so, if Working Memory is for a certain period of time dominated by the microassertions representing "**Not agree**", the individual will surely reject the proposal of having date. Of course, despite the determination of refusal, the subject may think after a while "What's a fool I was! I wish I had agreed" which will be explained in terms of meme interactions, or more precisely speaking, in terms of meme population dynamics.

A supply of memes to Working Memory and birth of new memes could quickly lead to a congestion and block of any meme movements. Hence, I have employed a mechanism facilitating a reduction of number of memes. When, say, "**Not agree**" meets "**Agree**", they annihilate leaving this way to places in Working Memory. The same applies to primary microassertions as, for example, "**rich**" and "**not rich**". Also the principle that younger meme eliminates older meme can be implemented. This kind of meme interaction can be called *combat*. In order to increase their penetration power, memes are forced to change direction of their motion. This kind of interaction can be called *collision*. Owing to collisions, each meme has a chance to visit any place in Working Memory and meet its mate.

Meme structure

Let us introduce the notation collected in the Table 4.4:

`R0` for "**rich**",	`r0` for "**not rich**",
`N0` for "**nice**",	`n0` for "**not nice**",
`A0` for "**Agree**",	`a0` for "**Not agree**",
`AN` for "**Agree if nice**",	`AR` for "**Agree if rich**".

Table 4.4. Notation for meme informational content

As it can be seen, every microassertion is denoted as a unified three-element sequence of characters. If only the second character is non-zero, this means that a given microassertion refers to a fact. Otherwise a given microassertion refers to a rule.

Using the above notation we can consider mating in terms of genetic crossover applied to informational entities. As it can be seen in the Fig. 4.2, the birth of new microassertion **A0** has been done via simple exchange of appropriate characters between two parent microassertions **AR** and **R0**. This means that even information can have its genetic-like code. The resulting microassertion **0RR** makes no sense, so it either turns back **0R0** or disappears from Working Memory.

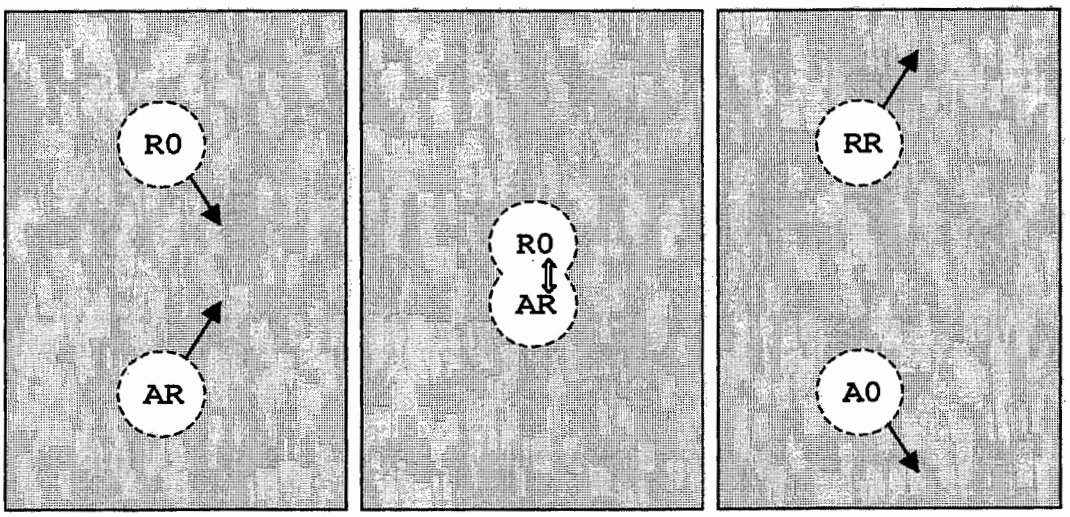

Fig. 4.2. Three snapshots showing the mechanism of meme mating. A micro-assertion **R0**, contributing to the subject's feeling that a date-proponent is rich, meets the microassertion **AR**, contributing to the subject's principle of dating only with somebody nice and rich, and as a consequence of a genetic-like crossover, the microassertion **A0** contributing to the subject's readiness to have the date is born.

Other discussed meme interactios has no their biological counterparts. However, an algorithm of character recombination is not too sophisticated in such cases as a crossover with negation (Fig. 4.3) or annihilation (Fig. 4.4). The fourth kind of interaction results in change of meme movement direction with preserving of the meme's informational contents. This takes place when the memes that met have nothing in common or are the same.

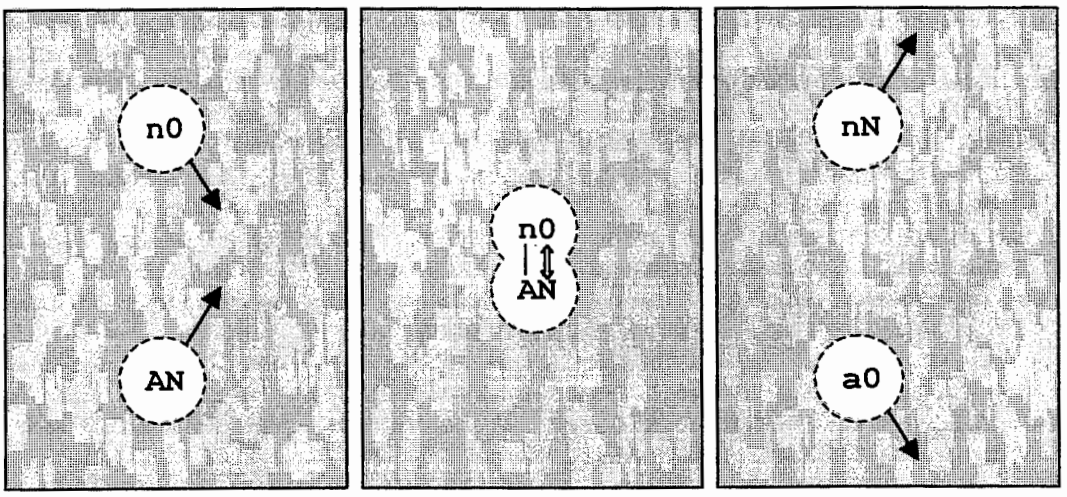

Fig. 4.3. The mechanism of meme mating with negative crossover. A microassertion **n0**, contributing to the subject's feeling that a date-proponent is not nice, meets the microassertion **AN**, contributing to the subject's principle of dating only with somebody nice and rich, and as a consequence of a negative crossover, **a0** is born. The microassertion **a0** contributes to the subject's determination to resign from the date. The **nN** must die since it has illegal syntax.

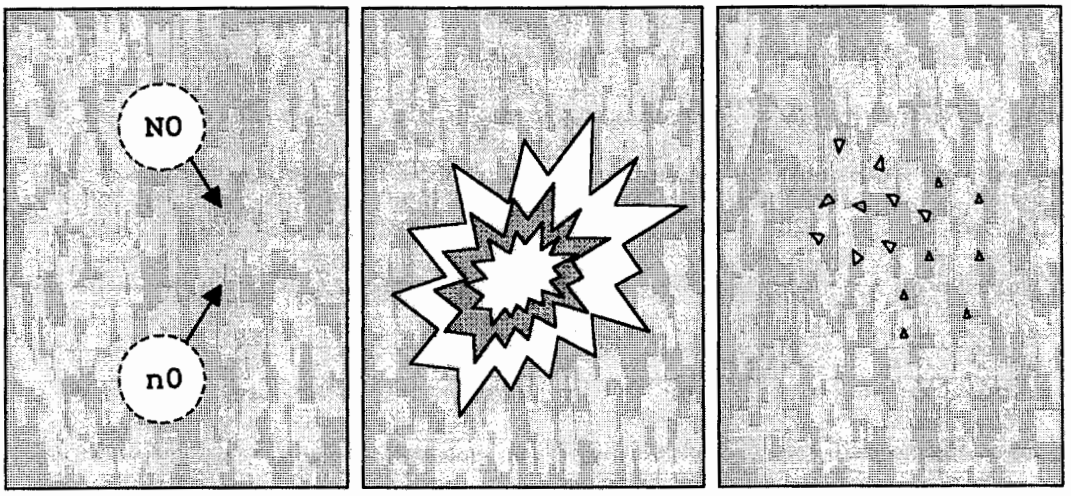

Fig. 4.4. Annihilation of contradictory memes. A microassertion **n0**, contributing to the subject's feeling that a date-proponent is not nice, meets the microassertion **N0**, contributing to the subject's feeling that a date-proponent is nice, and both dissapear from Working Memory.

Analogical set of interaction rules applies to memes of more sophisticated structures. Let us consider the situation when a subject perceives a ball for the first time and, at the same time, learns the name of the perceived object. In such a case, Working Memory starts admitting a stream of bimodal memes. Each of the bimodal memes contains a video track and an audio track (Fig. 4.5) and contributes to the knowledge that a small round object is associates with the sound effect one can produce reading the word 'ball'.

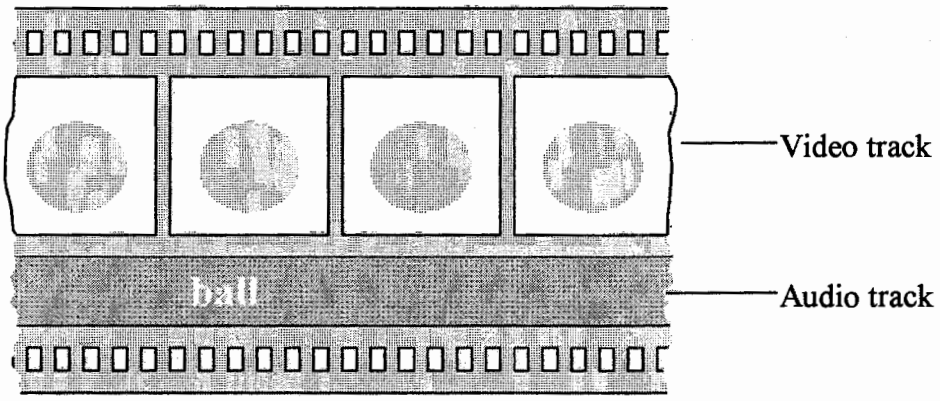

Fig. 4.5. A conceptual scheme of a bimodal meme. Its representation in the brain supposedly is a bundle of spiketrains produced by nerve cells connected to subject's sensorium.

Let us denote such a meme B_1. Since Working Memory is only a short-term store, some representatives of the population of B_2's find their asylums at appropriate compartments of Semantic Memory. Now let us assume that the after a time the subject saw a small round flying object while nobody commented this event. The subject's sensorium starts producing bimodal memes we will denote B_2. Every B_2 has empty audio track and contributes to the knowledge that a round shape has already been perceived. Since a number of B_2's reached the channels to Semantic Memory, it starts releasing related memes based on similarity of the contents of video tracks. The higher similarity the higher density of stream of copies of the meme in Semantic Memory selected to enter Working Memory. Let us assume that the only considerable stream coming from Semantic Memory contains memes of the type B_1's that contributes to the knowledge that a similar round shape can be associated with the sound 'ball'. Every encounter of the B_1 and B_2 results in a production of a copy of the meme B_3 being a kind of *subtraction* of B_2 and B_1 (Fig. 4.6). B_3 inherits only the content of audio track. Although B_2's video track and B_1's video track are not perfectly equal, B_3's video track becomes empty. When the population of B_3's becomes a dominating meme population in in

Working Memory, the subject will become aware that the perceived round shape is called ball.

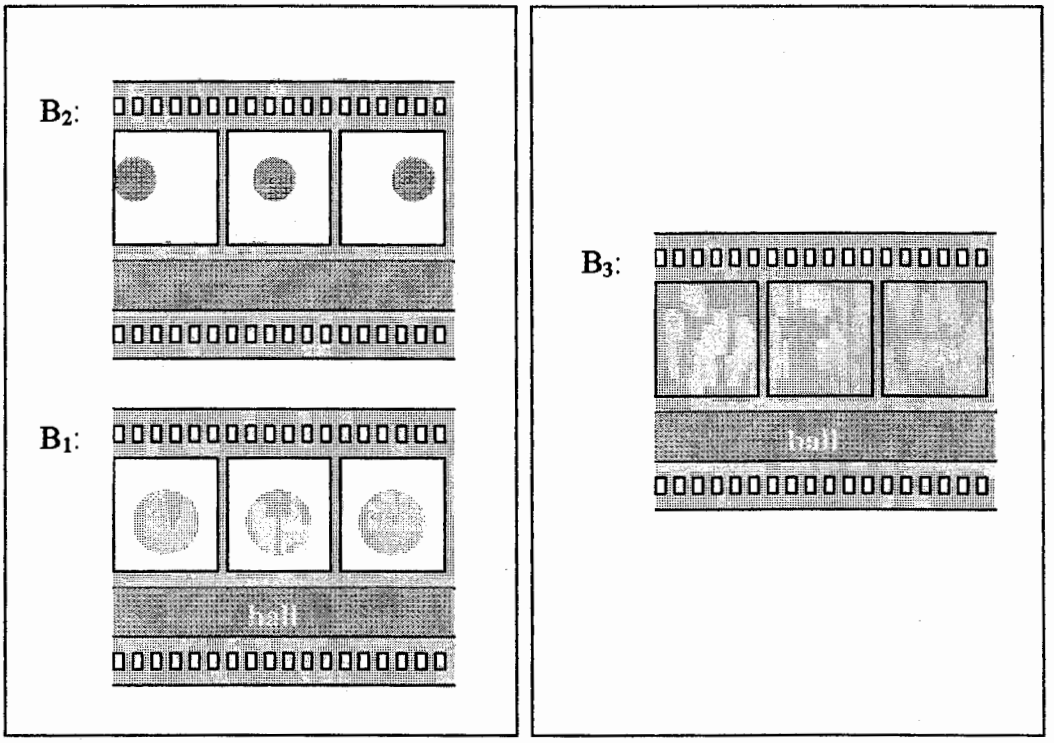

Fig. 4.6. Two snapshots of bimodal meme interaction. The meme B_2 contributes to the knowledge that a round shape has already been perceived. The population of B_1's comes from Semantic Memory and contributes to the knowledge that a similar round shape can be associated with the sound effect one can produce reading the word 'ball'. As a result of their encounter the meme B_3 is produced. The B_3 is a kind of *subtraction* of B_1 and B_2. When the population of B_3's dominates over Working Memory, the subject will become aware that the perceived shape is called ball.

Summary

As a smallest unit of meaningful information processed in mind I propose a kind of meme I christened *microassertion*. A microassertion alone means nothing. A population of microassertions carrying the same information "`X`", when dominates over Working Memory, can cause a subject's feeling or belief that "`X`". A microassertion's can have a two-element structure *HT*, where *H* is "head" and *T* is "tail". If the tail is empty, as for `XO`, the microassertion content equals the fact that "`X`". If the tail is not empty, as for `YX`, the microassertion content equals the rule that "`Y if X`". Microassertions can interact in the way resembling genetic crossover. For instance, a microassertion `XO` can mate the rule `YX`,

and, as a result of the union, a new fact Y0 is born. A single act of such an interaction has no significance, but if it takes place at the same time in a number of places in Working Memory, a population of the resulting fact Y0 has a chance to dominate over Working Memory and cause the subject's feeling or belief that "Y". If the facts "X" and "x" are contradictory, the encounter of the microassertions YX and x0 results in the birth of the microassertion y0 and that can contribute to the subject's feeling that "not Y", while the encounter of the microassertions X0 and x0 results in annihilation of both of them. The proposed mechanism facilitating the encounters consists on a continual migration of microassertions in the Working Memory space. As a result of a perception of an object or situation, Semantic Memory produces streams of memes carrying representations of categories the perceived object or situation belongs to. If a membership to a given category "X" is not obvious, two streams of memes—one containing copies of X0 and one containing copies of x0 (representing "not X") are produced and directed to Working Memory. A proportion of the streams' densities is a function of a membership of the percept to the fuzzy set X. The same applies to other related categories. Hence, the Working Memory becomes a war-theater where population of contradictory memes fight for domination and, as a consequence, for the subject's feeling. Based on the concept of multimodal microassertions a mechanism of language acquisition can be explained.

...Perhaps we are reluctant
to give up our claims for human uniqueness
—of being the only species that can think big thoughts.
Perhaps we have "known" so long that machines can't think
that only overwhelming evidence can change our belief.[64]

—HERBERT A. SIMON
Nobel Prize winner in economy

5. Neural "Lego blocks"

All meme interactions discussed in the previous chapter can be programmed in any of high-level symbolic languages and run on an ordinary computer. Unfortunately, such an approach is applicable only for isolated processes. Even the most powerful of programmable computers seem to be not sufficient for simulating a thinking brain as a whole. Hence, as an ultimate solution for a modeling of mind I propose a scalable hardware working as an imitation of a neural structure. Neurobiological plausibility of such model is a bonus. In this chapter I present an idea of artificial brain building based on standardized neural modules.

Tile-based memory

How to make memes navigate and interact? In the framework of the solution I propose Working Memory consists of standardized tiles, while memes, represented by limited-length spiketrains, navigate in the course of jumping from one tile to another. The tiles can be arranged in a countless number of ways, however, for practical reasons, regular arrangements are preferable.

Among regular tile arrangements we can first distinguish 1-dimensional, 2-dimensional, 3-dimensional (denoted 1D, 2D, 3D, respectively) and 4- or more-dimensional structures. In the last case, the structure is physically 3-dimensional, but describable in terms

[64] Simon (1995)

of 4- or more-dimensional geometry. The second essential measurement in a tile-based memory is a number of neighbors each tile is connected to. And so, we can talk about *m*-connected (denoted *m*C) memories, where *m* can equal 2, 3, 4 or more. It also can be essential whether the connections are unidirectional (denoted u) or bi-directional (denoted b). And so, if a tile-based memory is chessboard-like and each tile is connected only with neighbors of opposite color in bi-directional manner, we can denote it 2D4Cb, which means 2-dimensional, 4-connected, bi-directionally.

Any *n*D*m*C memory consisting of a limited number of tiles has its boundary. The border tiles have neighbor-free sides. Any of the neighbor-free sides can be connected to another neighbor-free side of the same tile. Such a solution, however, makes some irregularity in the memory model. A solution that preserves regularity is to arrange the tiles to make them forming a rectangular, cuboid, or hypercuboid, and connect border tiles with appropriate border tiles at the opposite side of the memory. This way a tile based memory becomes a sort of torus. In a toroidal memory a meme, if not takes part in a collision, may keep its direction for indefinite time returning periodically to the same place.

A task for each of the tiles is twofold. First, a tile is to admit incoming memes and, after a certain delay, to send them to certain neighbor tiles. Second, a tile is to facilitate such meme interactions as crossover, combat, and annihilation. As it can be seen, a tile-based memory has all features defining cellular automata. For the sake of biological plausibility the tiles are to behave as simplified neural networks. As for the neural networks, they can be grown in a lower-degree cellular automata space. Hence, the proposed implementation of the "4 + 1" memory model is a 2-degree cellular automata device, where the higher-degree cells are called tiles.

Figure 5.1 shows a toroidal 2D6Cu model of Working Memory called *MemeStorms1* or *MS1* in which the tiles had hexagonal shapes and could receive memes from a continuous string of three different neighbors and send memes to other three neighbors. As it can be seen, in the MS1 the directions of all meme velocity vectors are contained in the range from -60^0 to 60^0. This means that the memory architecture reinforces a sort of unidirectional meme stream that in the diagram in Fig. 5.1 flows bottom-up.

A complete model of Working Memory contains input points (used for injection of new memes into Working Memory), as well as output points (used for memory state recognition). For the sake of clarity, neither input nor output points are shown in the Fig. 5.1.

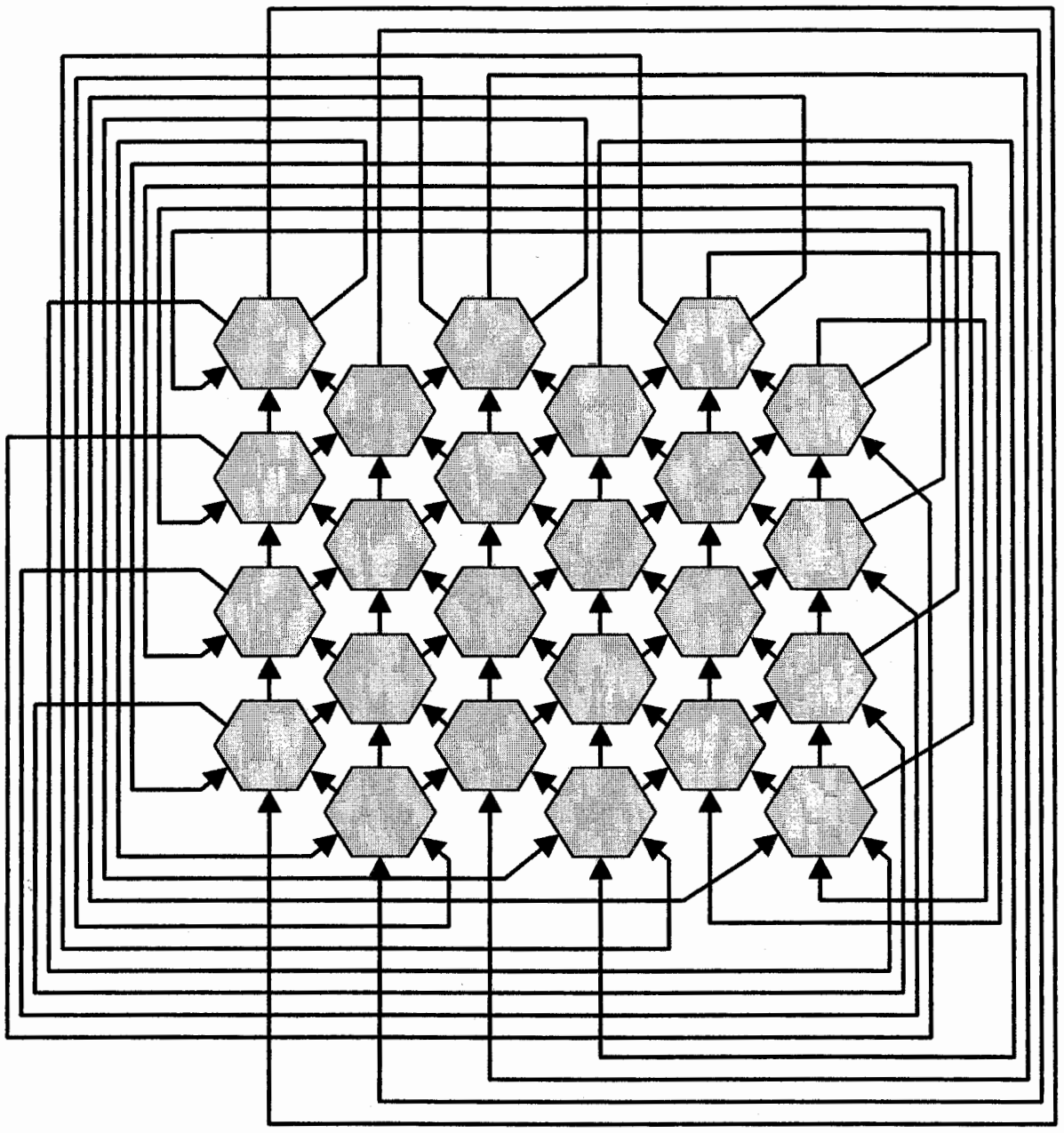

Fig. 5.1. A toroidal 2D6Cu tile-based Working Memory *MemeStorms1* (*MS1*)

Tile-based dynamics

The described model of Working Memory (WM) when a stream of memes is provided to its input points becomes a discrete dynamical system. Let S be a stream of memes entering WM. As a control parameter we can take the set $\{ (M_i, d_i) \mid i = 1, ..., n \}$, where M_i is an unique meme, d_i is an density coefficient for copies of memes M_i in S, n is the number of different kinds of memes in S. If the density of S is assumed to equal 1, the average density of the stream of copies of M_i equals $d_i / (d_1 + ... + d_n)$. The WM's state at the time t is a location, velocity vector and membership of all memes roaming there in the time t. The state defined in the above way means itself nothing. What can be meaningful is a particular derivative of the state called order parameter.

There is practically unlimited number of possibilities of defining order parameters. It may be existence vs. non-existence of a particular constellation of memes in Working Memory. It may be a coincidence of their velocity vectors. Let us consider one of the simplest possible, but strongly meaningful, order parameters: the h_M factor calculated as $N_M/(N_M + N_{\sim M})$, where M is a meme of interest and N_M is actual number of copies of the memes M, while $N_{\sim M}$ is actual number of copies of the meme ~M, where M and ~M are contradictory memes. A history of subject's belief or feeling M vs. ~M is a function of a history of changing values of the h_M factor. Figure 5.2 shows sample plots of h_M for three different values of control parameters, $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, taken during an experiment on a certain model of Working Memory of the class 2D6Cb (see Figure 5.1) applied to three versions of the story from the Table 4.2. The taken values of control parameters were:

$\mathbf{a}_1$ = { (N0, 16), (n0, 24), (R0, 16), (r0, 24), (AN, 14), (AR, 14), (00, 0) },
$\mathbf{a}_2$ = { (N0, 16), (n0, 24), (R0, 24), (r0, 16), (AN, 14), (AR, 14), (00, 0) },
$\mathbf{a}_3$ = { (N0, 24), (n0, 16), (R0, 24), (r0, 16), (AN, 14), (AR, 14), (00, 0) },

The control parameter $\mathbf{a}_1$ reflects the situation where the subject's certainty that the perceived person is **nice** is 40% (and that **not nice** - 60%), while the certainty that the perceived person is **rich** is also 40% (and that **not rich** - 60%). The equal density coefficients for the memes **Agree if nice** and **Agree if rich** reflect the equal significance of niceness and richness. For such distribution of density coefficients we should expect quick victory of population of memes telling **Don't agree!** over the population of memes telling **Agree!**

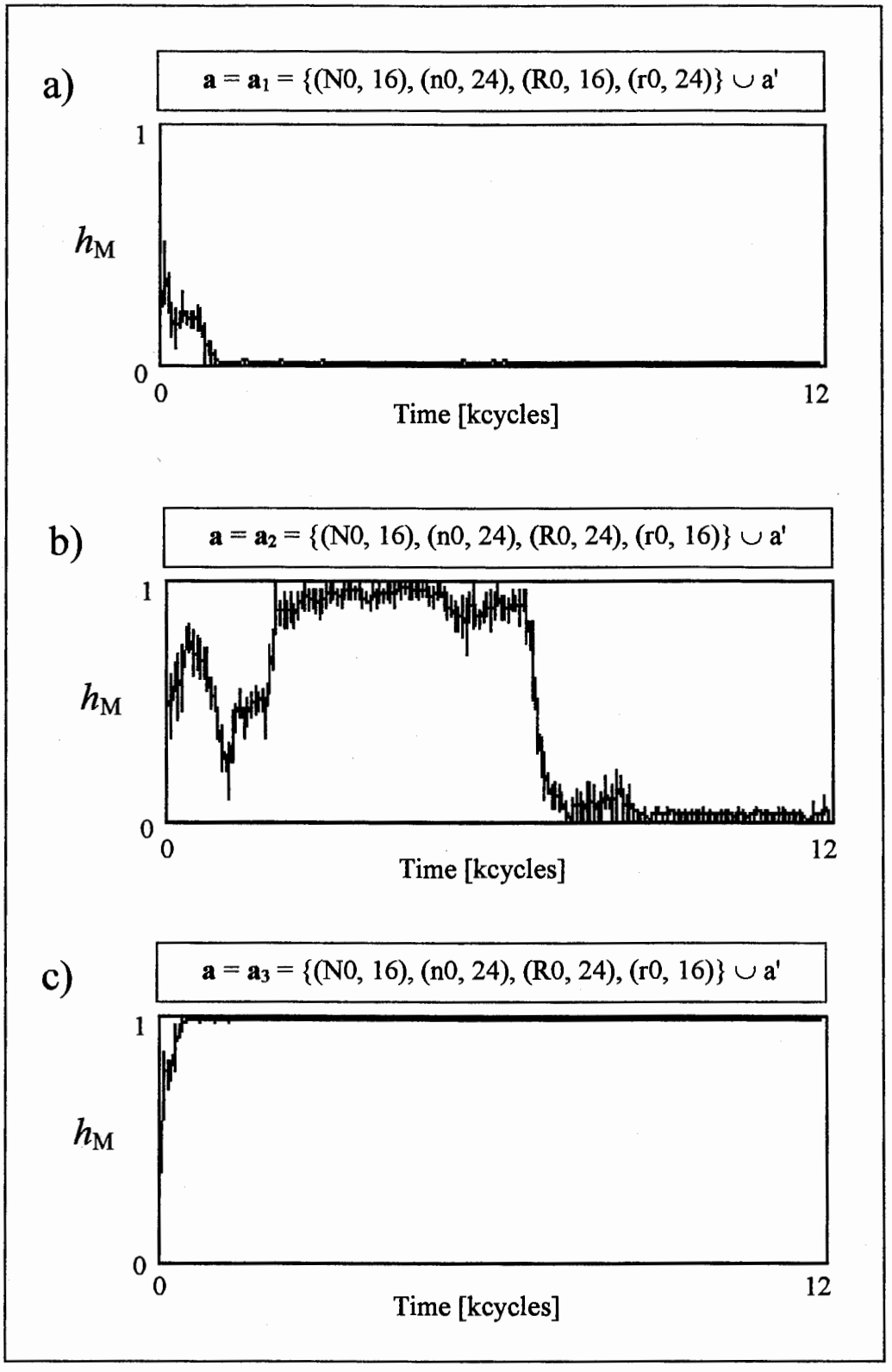

Figure 5.2 Dynamics of the state of simulated Working Memory WM). The WM's size is 20x16 hexagonal tiles. The state, represented by the order parameter h_M, can change over time in several ways, depending on the value of control parameter **a**, where a' = {(AN, 14), (AR, 14), (00, 0)}. The plot obtained for $\mathbf{a} = \mathbf{a}_2$ strongly resembles empirical feelings expressed by a human subject using a mouse (see Figure. 2.1B).

The control parameter $\mathbf{a_3}$ reflects the situation where the subject's certainty that the perceived person is **nice** is 60% (and that **not nice** - 40%), while the certainty that the perceived person is **rich** is also 60% (and that **not rich** - 40%). The significance of niceness and richness is equal. For such distribution of density coefficients we should expect quick victory of population of memes telling **Agree!** over the population of memes telling **Don't agree!** (provided no better candidate is considered at the same time by the subject).

The control parameter $\mathbf{a_2}$ reflects the situation where the subject's certainty that the perceived person is **nice** is 40% (and that **not nice** - 60%), while the certainty that the perceived person is **rich** is 60% (and that **not rich** - 40%). The significance of niceness and richness is equal. What should we expect for such distribution of density coefficients? There are no grounds for quick victory of population of memes telling **Agree!** over the population of memes telling **Don't agree!** or vice versa. One expectation could be that h_M for M=**Agree** will permanently close 50%.

As in can be seen in the Figure 5.2, in case of $\mathbf{a_1}$ or $\mathbf{a_3}$ the value of the order parameter h_M goes quickly to the nearest neighborhood of an appropriate stable point (0 or 1, respectively). In case of $\mathbf{a_2}$, the value of the order parameter oscillates sometimes near 0 and sometimes near 1 in the way suggesting the occurrence of an attractor having two unstable foci. Does not the plot of for resemble one of the plots taken when human subject's expressed their feelings using a computer mouse (see Figure 2.1B)?

The results of a more careful investigation of the dynamics of tile-based models will be discussed in the next chapter.

From order parameter to a belief

As it can be learned from the simulation results shown in the Fig. 5.2, a given population of memes can take or loss a domination over the Working Memory for different periods of time. Sometimes the periods are very short, so, for the sake of psychological plausibility they should not cause immediate change of subject's belief. Hence, a kind of inertial filter has been employed. The filter, for a given plot of $h_M(t)$ produces the plot of $E_M(t)$ such that

$$E_M(t) = h_M(t) + kE_M(t-1) - 0.5$$

where k is a coefficient expressing a strength of influence of previous values of E_M to the current value of E_M. The plot of the function E_M much stronger than h_M taken based on the h_M

shown in the Figure 5.2.b for k=0.9 resembles the trajectory of evaluation taken using the Mouse Paradigm (see Figure 2.1.B).

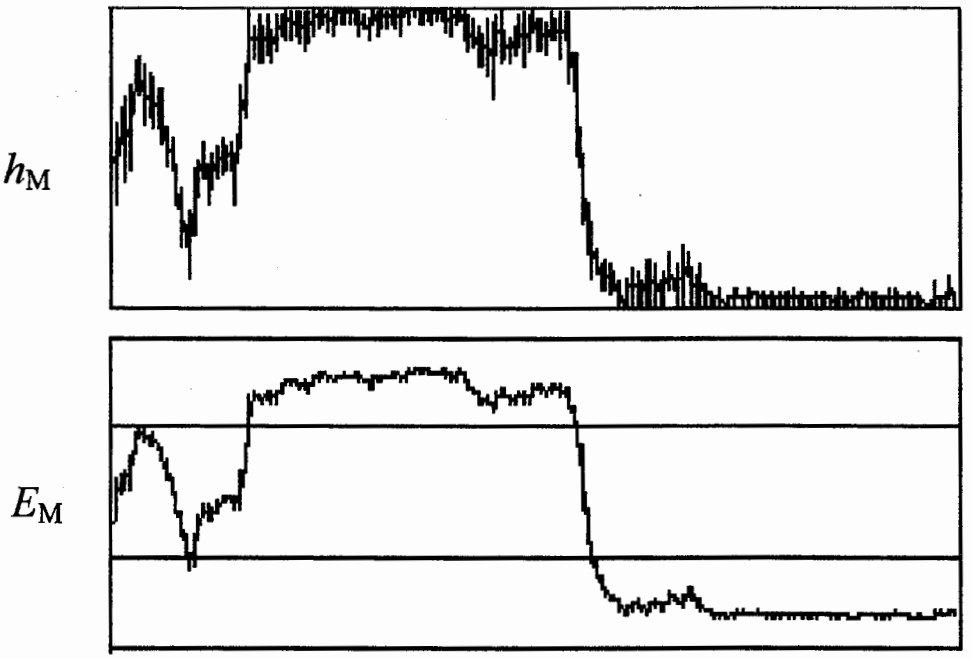

Figure 5.3. The plot of h_M (order parameter of the modeled Working Memory) and the plot of E_M to be interpreted as an evaluation trajectory reflecting feelings about a person or social situation. E_M is produced by applying of an inertial filter to h_M.

What happens when a subject must verbalize a decision based actual feelings, especially when the decision is of the kind "yes/not"? In the proposed model the choice "yes" or "not" depends on the value of another state variable reflecting decision readiness.
Let us introduce a binary function J, such that

$$J(t) = M \textbf{ if } (B_M(t) > -\varepsilon \textbf{ and } J(t-1) = M) \textbf{ or } (B_M(t) > \varepsilon \textbf{ and } J(t-1) = \sim M),$$
$$J(t) = \sim M \textbf{ if } (B_M(t) < \varepsilon \textbf{ and } J(t-1) = \sim M) \textbf{ or } (B_M(t) > -\varepsilon \textbf{ and } J(t-1) = M),$$

where ε and $-\varepsilon$ are certain critical values of causing a change of the value of J. This means, that if the subject is to provide his/her decision about the date, the decision may depend on the time that passed since the problem had been considered. As it can be seen in the Figure

5.4, despite the chaotic-like, frequent changes of feelings E_M, the changes of the subject's readiness for a particular decision are relatively much less frequent.

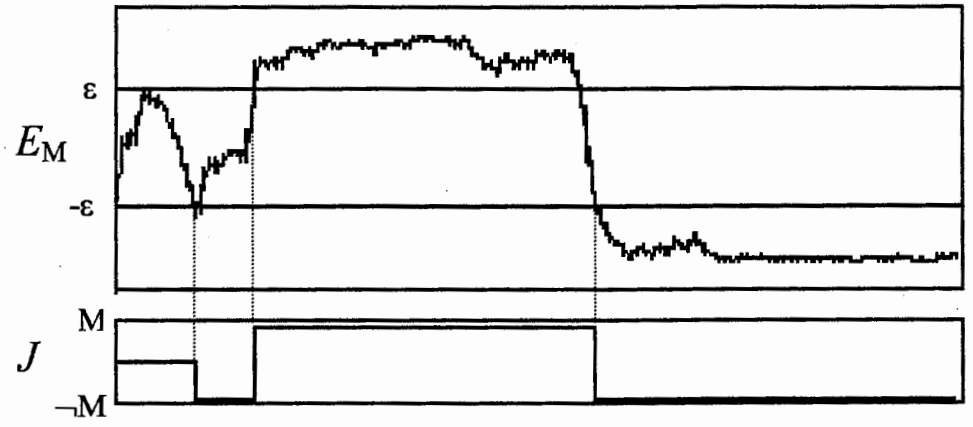

Figure 5.4. Subject's readiness to express a particular decision M or not M based on the history of his/her feelings E_M. The critical values ε and -ε could be interpreted in terms of individual differences. Impulsive persons have small ε.

Inside a tile

Each of the tiles constituting Working Memory consists of appropriately connected subtiles, while each of the subtile consists of appropriately connected circuits consisting of appropriately connected subcircuits. On the lowest level of the hierarchy there are neurons. Figures 5.5 to 5.9 show the upper part of the modular hierarchy (without the neural level) designed for the *MemeStorms1* Working Memory model.

According to the *MemeStorms1* meme processing principles the Working Memory is a instance of a 2D6Cu grid, where each tile has three input channels and three output channels. So, on the first level of division, a tile consists of three subtiles, where each of the subtiles has three inputs and a single input equal to one of the tile's outputs (Fig. 5.5). A subtile's task is to receive all memes entering the tile in a given moment and, if applicable, to produce a resulting meme to leave the tile via an appropriate output channel. Since it has been assumed that simultaneous arrival of three memes to a tile causes no interaction, one of the elements of the subtile is a "memetic XOR" that lets a coming meme pass through if and

only if no other meme comes at the same moment. The second element of the subtile is an Interactor facilitating all changes in a given meme's structure resulted from any of the assumed kinds of interactions (Fig. 5.6). The internal structure of the Interactor is shown in the Figures 5.7 to 5.9.

Following the way a pattern passes the internal structure of a single tile one can see that the total time a pattern must spend inside the tile is close to 10/f, where f is the maximum frequency of neuron firing. Hence, if the model were build using biological neurons ($f \approx 1000$ khertz), the time would equal about 10msec. This means that the length of simulation that equaled 12000 kcycles can be interpreted as an equivalent of 120 sec., i.e. the length of Vallacher and Nowak's experiments with human subjects

Conclusions

The proposed technical solution of meme migration and interaction in Working Memory is based on cellular automata paradigm. The Working Memory is build as a grid of tiles, and meme can navigate jumping from a tile to another tile. Several Working Memory structures have been investigated. The most interesting results were provided by a model called *MemeStorms1* that consists of hexagonal tiles supporting three-direction meme traffic. When streams of contradictory memes were directed to the model, for certain distributions of stream densities it has been noted that population of resulting contradictory memes could dominate by turns over Working Memory. This phenomenon strongly resembles oscillation of the value of an order paramemter of a dynamic system. Hence, a definition of an order paramemter as a single real number representing a certain aspect of a state of Working memory has been formulated. Also formulas transforming a plot of changing order parameter over time onto a history of subject's feelings and a history of readiness to express a particular decision has been proposed. As for the tile internal structure, a hierarchy, where a tile divides onto subtiles, a subtile divides onto circuits, a circuit divides onto subcircuits, and a subcircuit is a network of appropriately interconnected neurons has been designed. An analysis of the neural structure shows that if the modeled neurons worked with the same frequency as biological neurons work, the time of meme processing in a single tile would be about 10msec. This means that the 12000 kcycle simulation can be interpreted as an equivalent of 120 sec., i.e. the length of Vallacher and Nowak's experiments with the mouse paradigm. This coincidence strongly supports the hypothesis about the biological plausibility of the discussed model.

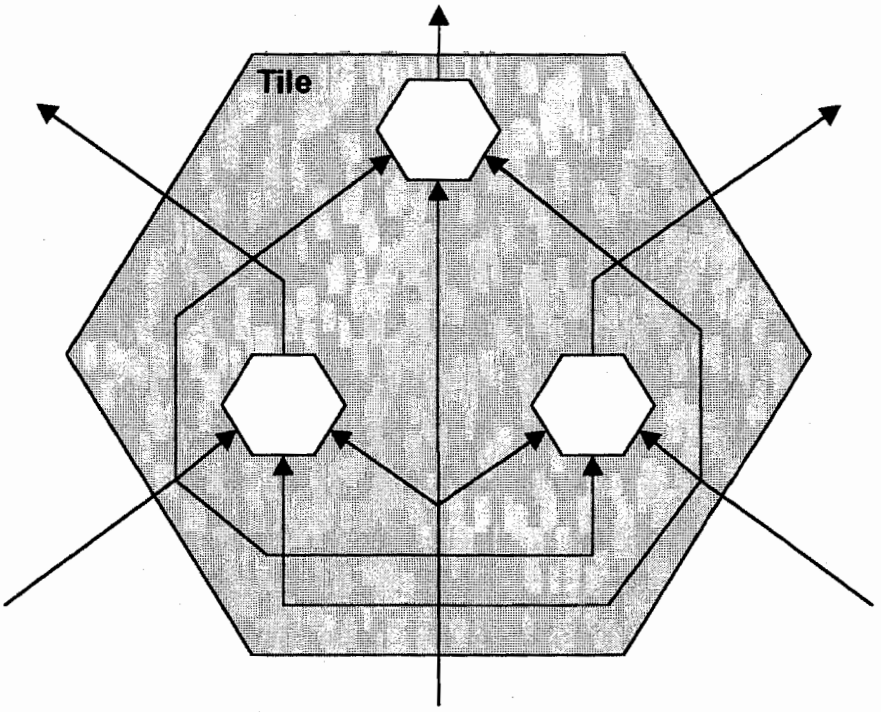

Figure 5.5. Tile internal structure. In the proposed solution a tile consists of three identical sub-tiles appropriately connected.

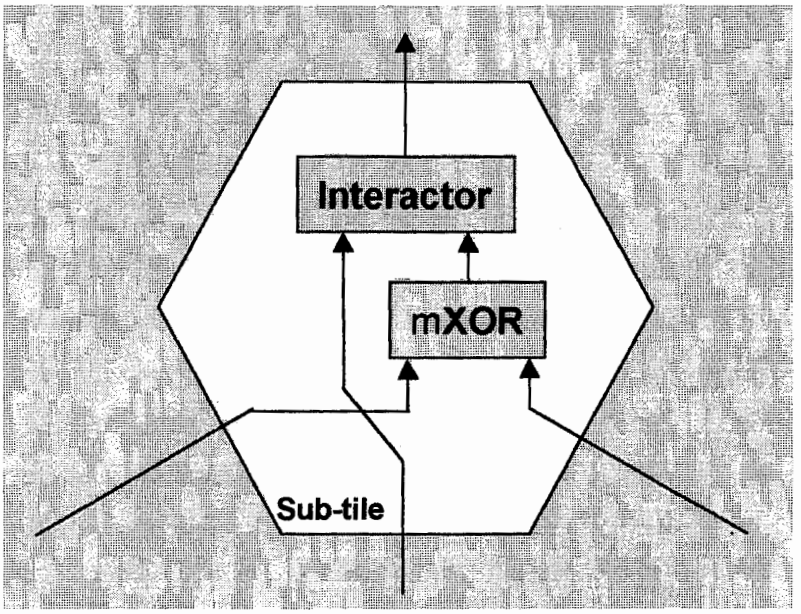

Figure 5.6. Sub-tile internal structure. Interactor produces a conclusion form two memes. The circuit mXOR (memetic Exclusive OR) returns meme x if and only if x came to one of its inputs while nothing came to the second input.

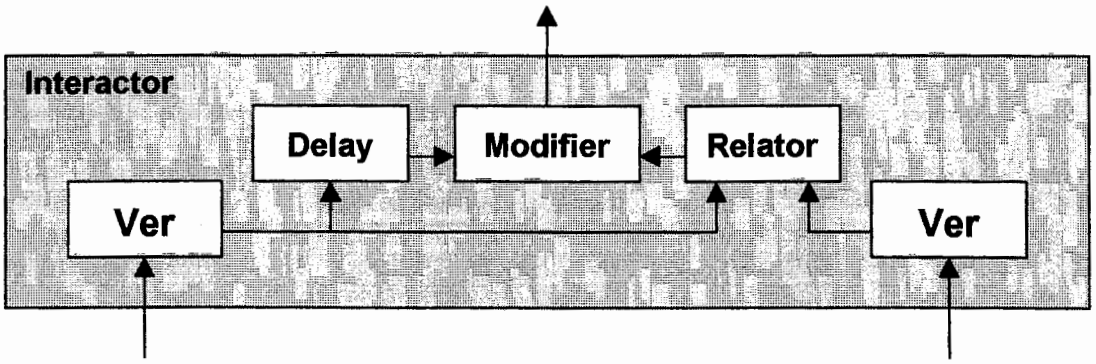

Figure 5.7. Interactor internal structure. When two incoming memes have proper syntax (to be checked by the circuit Ver), based on their identified relationship (to be done by the circuit Relator), the meme that came through the left inlet is appropriately modified.

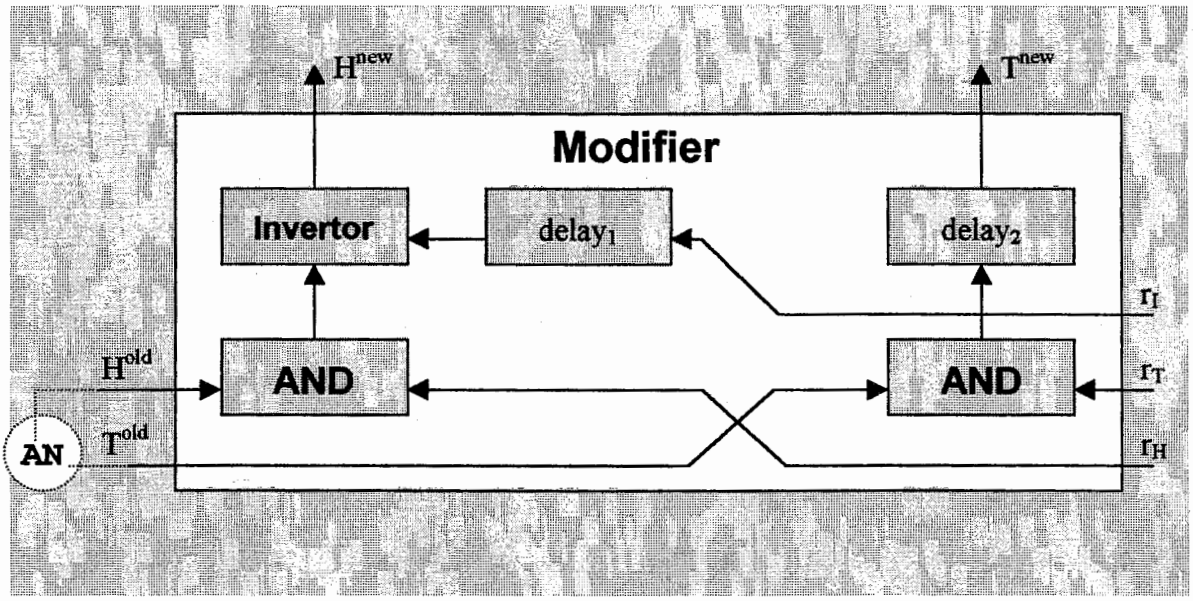

Figure 5.8. Modifier internal structure. A meme **AN**, taken as an example, depending on the values of r_H, r_T and r_I that encode the relationship with an interacting meme, may remain unchanged or turn to either **A0** or **a0**. A meme **A0**, if was taken as another example, depending on the values of r_H, r_T and r_I, might remain unchanged or disappear.

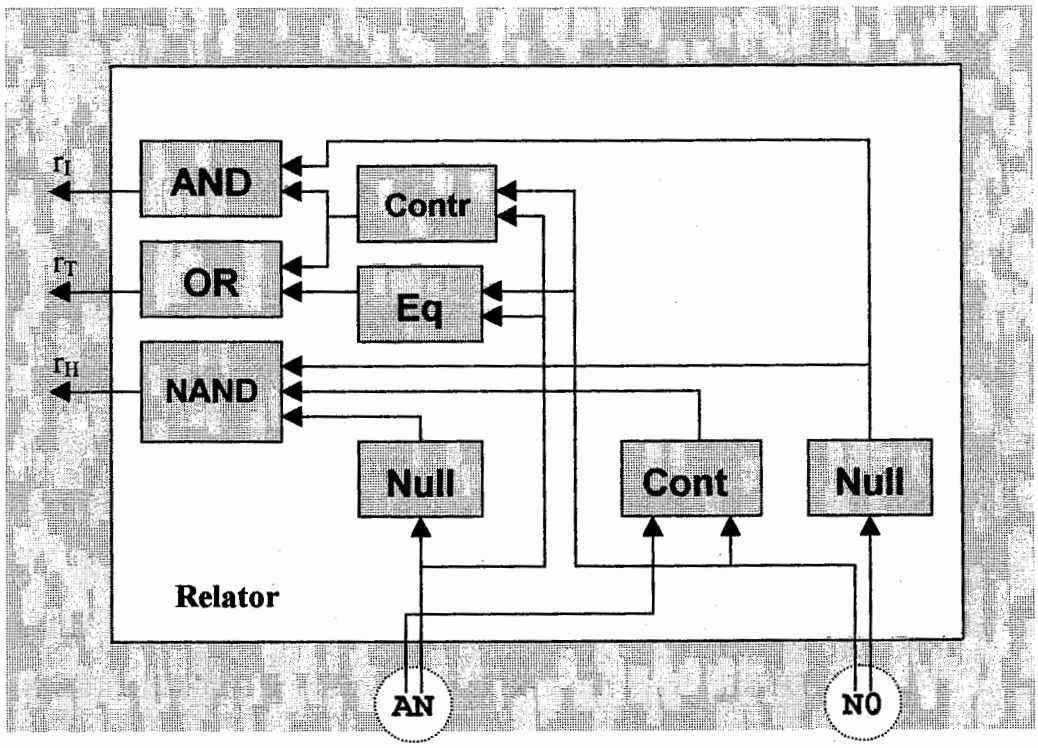

Fig. 5.9. Relator internal structure. The subcircuit **Null** returns True if the incoming part of a meme is **0**. The subcircuit **Contr** returns True if two incoming meme parts are contradictory. The subcircuit **Eq** returns True if two incoming meme parts are identical. AND, NOR and NAND are elementary logical gates. Owing to the system of interconnections the Relator recognizes a kind of interaction the two incoming memes can have and codes it as an unique value of the vector (r_H, r_T, r_I). In case of the memes **AN** and **NO** *(to mate with positive conclusion)*, taken as an example, the output vector will be (1, 0, 0). For **AN** and **nO** *(to mate with a negative conclusion)* it would be (1, 0, 1). For **AO** and **aO** *(to annihilate)* it would be (0, 0, 0). For **AN** and **RO** *(to collide elastically)* it would be (1, 1, 0).

The thought behind I strove to join
Unto the thought before,
But sequence raveled out of reach
Like balls upon a floor.

—EMILY DICKINSON
from *The Lost Thought*

6. Simulated streams of 'thought'

This chapter presents some additional experiments with the MemeStorms model aimed to confirm its psychological plausibility. The investigated phenomenon is an oscillation of social judgment in absence of new data about a perceived object or situation. The oscillation should occur more often in case of ambiguous cues, while in case of clear cues towards a particular judgment the oscillations should be seldom or not occur at all.

Experimental assumptions

A set of N=20 individuals has been taken from potentially existing population of all possible models endowed with a toroidal tile-based working memories of the 2D6Cu class. Each of the individuals had tiles configured as a grid of 20 x 16 tiles. Each of the individuals perceived a set of cues that were transformed onto four streams of memes. There were also two streams of memes carrying related rules producing memes contributing to two contradictory judgments. As a result of meme interactions a given individual expressed particular feelings. The judgment was expected to change over time despite constant average densities of incoming meme-streams. The only individual differences consisted in different schedules of meme supply, which was to reflect individual differences in neural circuitry. In face of known psychological evidence, if the model were psychologically plausible, the judgment should sometimes switch from a highly positive/negative value to an opposite value and the changes should occur more often in case of ambiguous cues, while in case of

expressive cues they should be more seldom or not occur at all. In order to make the experiment comparable with the data taken by Vallcher and Nowak (see Figure 2.1) the duration of scanning of "feelings" was taken 120 seconds (represented by 12 kcycles) for each of the individuals.

Experimental variables

Let us take a cue ambiguity as an independent variable. It will be a real number calculated as a function of four numbers representing densities of meme streams entering an a given individual's Working Memory. In the test-bed example there are two pairs of streams of contradictory memes: a stream of **N**0s (fact **N**), a stream n0s (fact **not N**), a stream of **R**0s (fact **R**), and a stream of **r**0s (fact **not R**)[65]. Let densities of the streams are denoted d_N, d_n, d_R, and d_r, respectively. The cue ambiguity factor a_C will be a function of the four densities. For the presented experiment it was assumed that: $a_C = 1 - | d_N / (d_N, + d_n) + d_R / (d_R, + d_r) - 1 |$. The cue ambiguity factor equals zero when either only two non-zero external streams of memes arrive to Working Memory and they are the stream of **N**0s with or **R**0s, or the stream of n0s with the stream of **r**0s. The cue ambiguity factor equals 1 when, for example, the number of incoming **N**0s equals the number of incoming n0s and the number of incoming **R**0s equals the number of incoming **r**0s.

The dependent variable is assumed to be a function of the frequency of individual's "changes of mind". For the presented experiment the frequency is calculated as Z = N(**A** → **a**) + N(**a** → **A**), where $N(j_1 \rightarrow j_2)$ is a number switchings from a readiness to make the decission j_1 to the readiness to make the decsion j_2 registered within a fixed period of time.

Results

In the first step of the experiment two resulting series of values of Z has been obtained, one series for

a = { (**N**0, 16), (n0, 24), (**R**0, 24), (**r**0, 16), (**AN**, 14), (**AR**, 14), (00, 0) },

the second series for

a = { (**N**0, 24), (n0, 16), (**R**0, 16), (**r**0, 24), (**AN**, 14), (**AR**, 14), (00, 0) }.

[65] The meaning of the symbols **N**, **R**, **n**, **r**, and **0** is explained in the Table 4.4.

which gives $a_C = 1$. The value of the last time clock was set as 12 kcycles (120 sec.). The consecutive values of Z were:

Series #1: 1 1 1 1 2 0 0 0 1 3 1 0 0 2 1 1 1 3 0 2 (average 1.05)

Series #2: 3 1 3 1 3 1 1 0 0 4 2 2 0 3 0 0 2 2 3 2 (average 1.65)

The difference between the average values of Z is counterintuitive since the symmetry of the sets of input data. Nevertheless, the t-Student test showed that the differences are not significant ($t = 1.167$, $p>0.05$). Also counterintuitive is the number of cases when Z=0, which suggest that often even in case of maximum cue ambiguity the model's feelings stuck in a highly positive or negative value and never turn the opposite value. In order to make confirm this supposition the simulation of one of such cases was repeated with the plot of h_M, E_M and J taken for the time from 0 to 150 kcycles (0 to 1500 sec.). The result showed that even seemingly steady feelings may unexpectedly turn opposite after a long period of time.

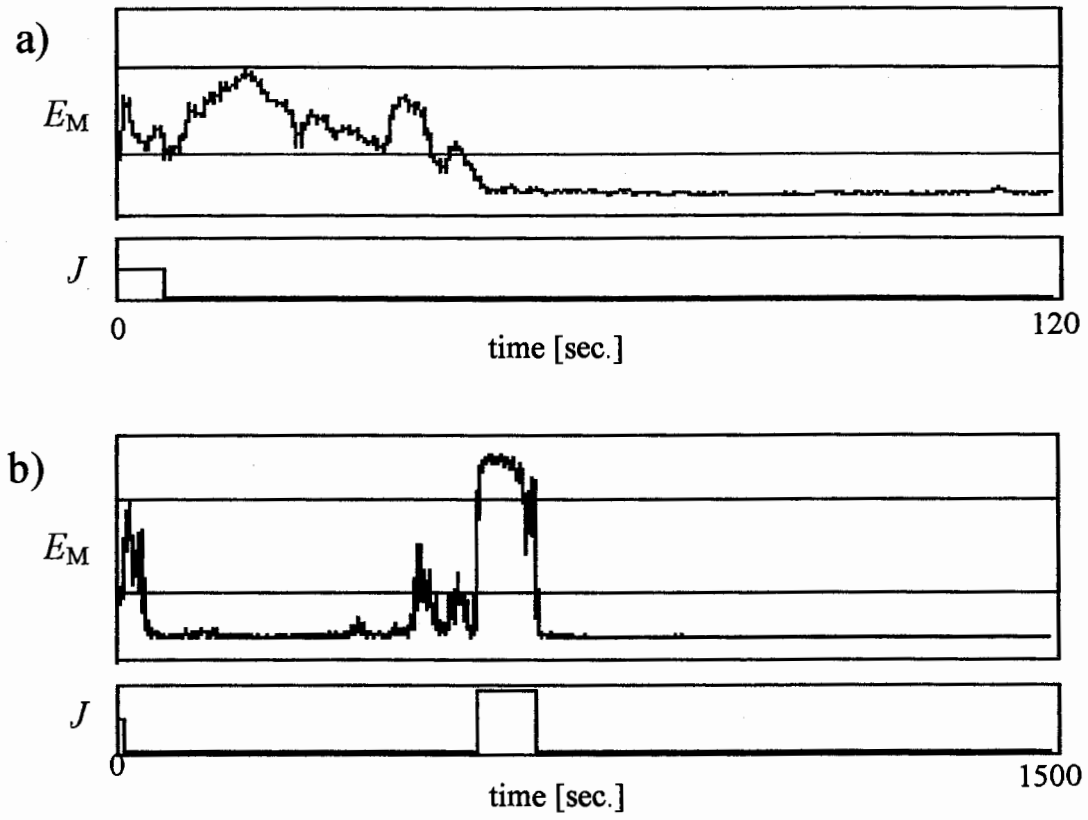

Figure 6.1. a) Non-oscillating feelings despite ambiguous cues (Z=0).
b) Longer duration of simulation of the same individual with the same initial data reveals oscillations (Z=3) that did not occured within the period of 120 sec.

In the second step of the experiment two resulting series of values of Z has been obtained, one series for

$$\mathbf{a} = \{ (\mathtt{N0}, 10), (\mathtt{n0}, 30), (\mathtt{R0}, 10), (\mathtt{r0}, 30), (\mathtt{AN}, 14), (\mathtt{AR}, 14), (\mathtt{00}, 0) \},$$

the second series for

$$\mathbf{a} = \{ (\mathtt{N0}, 30), (\mathtt{n0}, 10), (\mathtt{R0}, 30), (\mathtt{r0}, 10), (\mathtt{AN}, 14), (\mathtt{AR}, 14), (\mathtt{00}, 0) \}.$$

which in both cases gives $a_C = 0.5$. The value of the last time clock was set as 12 kcycles (120 sec.). The consecutive values of Z were:

Series #3: 0 0 0 0 0 0 0 (average 0.00)

Series #4: 0 0 0 0 0 0 0 (average 0.00)

The difference between average values of Z proved to be statistically significant for the series #1 and #3/#4 ($t = 2.904, p < 0.01$), as well as for #2 and #3/#4 ($t = 3.399, p < 0.01$).

This result confirms the thesis that the proposed model of Working Memory facilitates such a way of processing of perceived data, that modeled subject's feelings about a person or a social situation are, as in case of human subject, more stable when perceived cues are less ambiguous, while the bigger ambiguity of the cues, the higher probability that the feelings will change or oscillate. The comparison of the plots of changing value of order parameter formulated for the neural dynamic system representing modeled subject's working memory with the plots taken during the Mouse Paradigm-based experiments on human subjects strongly suggest that in both cases we observe the same sort of information processing.

If a work requires a big stuff of thinking,
If it can't be completed without others' support,
Did a thousand of madmen attempt it before?
Want to have clear conscience? Abandon such work.

—ADAM MICKIEWICZ
"A work"

Final conclusions

Since a thinking brain is commonly recognized as the most complex and the most perfect structure existing in the Universe, building a thinking machine seems to require the much more intellectual work than in the case of any device people ever built. It seems impossible that a single engineer could design all necessary circuits for an artificial brain. A number of projects aimed to create at least an illusion of machine thinking have been launched, but nothing impressive has been achieved in the framework of them. A number of experts maintain that a thinking machine cannot be built at all. Despite this, the idea of building of an artificial brain is far from being abandoned and the mad "brain-builders" are financially supported. Why?

First, as Artur C. Clarke noted, *when a scientist states something is impossible, he is very probably wrong*. Indeed, there was seemingly nothing more ridiculous in science as Simon Newcomb's mathematical "proof" that machines cannot fly. Second, in the field of artificial intelligence nobody expect impressive results soon. History of aviation also is long and full of dramatic faults. But every fault contributes to another trial that always has a good chance to be the first successful one. Third, the problem of the unimaginable amount of intellectual work can be eliminated in the course of an employment of the self-organizing powers that have been observed in chemistry, biology, and virtual realities. The powers result seemingly from an universal low of nature. Fourth, going to a great target one can, often by accident, discover something unexpected, however strongly contributing to our knowledge about surrounding reality.

The goal of the presented research is a possibly full computational model of human mind. The research has not been completed yet because they could not. Nevertheless, the results achieved till now seem to deserve to be presented in the form of a written work. The main result is a memory model called "4 + 1" and its simulated performance suggesting both psychological and biological plausibility. The concept of the model has been described, while some of its mechanisms has been simulated and compared with selected psychological evidences. In the "4 + 1" model five kinds of memories are distinguished: Procedural Memory, Filtering Memory, Episodic Memory, Semantic Memory and Working Memory. I myself designed and wrote a prototype of the program simulating the Working Memory. A program simulating a simplified model of the Semantic Memory has been written by Ms. Alicja Matusewicz in the framework of her M.Sc. work I supervised[66]. All of the theoretical and computational investigations that have been done on the "4 + 1" model suggest the following conclusions:

1. The "4 + 1" model is modular and dynamic, while modularity and dynamics are widely accepted in Cognitive Sciences community as mind's properties. The model also admits that the phenomenon of self-organization, that concerns several aspects of reality investigated in the framework of several scientific disciplines, must take place in mind.
2. The "4 + 1" model develops the SPI model of organization of memory proposed by Endel Tulving. The new contribution is the proposed diagram in which the Working Memory plays an integrative role, as well as a detailed model of mechanisms of Working Memory based on neural computation.
3. Although the proposed mechanism of Working Memory has been inspired by Marvin Minsky's *Society of Mind* and William Calvin's *Cerebral Code*, the developed solution based on the idea of a physical space inhabited by populations of navigating and interacting memes is original. A single meme has a minute significance. Only a population of given memes that dominated Working Memory is a subject's current knowledge, opinion, judgment or desire.
4. The proposed mechanism of meme interaction consists in a partial exchange of informational content in the way analogous to genetic crossover. This way the evolutionary mechanism becomes more universal since it seems to apply to reproduction of both biological entities and mental entities.

[66] Matusewicz (1999).

5. The "4 + 1" model explains a wide spectrum of psychological phenomena. It shows which way logical conclusions can be derived consciously or unconsciously. It shows which way great streams of perceived information can be integrated. It shows a way mind can cope with incomplete or contradictory cues. The idea of populations of navigating memes eliminates the problem known as conflict resolution (a choice of a pair of statements to be processed) that is significant in the production systems employing the first order logic (e.g. John Anderson's ACT-R[67]). Here every statements exists in a number of copies, while a number of acts of inferencing takes place simultaneously in several regions of Working Memory. Owing to this a population-based inferencing from contradictory statements becomes simple and natural. Finally, the "4 + 1" model explains the psychological evidence of oscillation of subject's feelings about a person or social situation even if perceived data remain unchanged.

6. During the simulation of processes in Working Memory it was observed that in some cases even when incoming streams of memes representing contradictory cues about a perceived person had constant densities, the populations of resulting contradictory judgments alternately managed to dominate. This property of the "4 + 1" model can be explained as a tendency of the state of the memory to go to a strange attractor. Hence, a dramatic change of subject's opinion can take place without any external reason as a sort of the Butterfly Effect. It was shown and statistically verified that in the absence of contradictory cues the related population of memes representing a judgment takes quickly a permanent domination over Working Memory.

7. While the two last conclusions suggest a strong explanational power of the "4 + 1" model, its empirical verification is difficult. One fact supporting the suggestion about the model's psychological plausibility is the qualitative resemblance between selected plots of human feelings scanned by Vallacher and Nowak using the Mouse Paradigm and selected plots generated by the model. Since the dynamics of human subject's feelings and the dynamics of the model's "feelings" can share the same time-scale, which results from the calculation of the model's speed with the assumed frequency of neuron firings the same as for biological neurons, we have a considerable support for the belief about the model's biological plausibility.

---

[67] Anderson (1993); Anderson & Lebiere (1988).

8. As for the order parameter of the Working Memory considered as a complex dynamic system, only the simplest possibility--the number of particular kind of memes versus other kind of memes--has been explored. In the future several kinds of other order parameters, including meaningful constellations of homogenous memes, could be investigated.

9. As for anticipations resulted from the model, first I would suggest that when the technique of cortical activity imaging reaches an appropriate level, we will able to see memes as moving constellations of neural firings. Other anticipation resulting from the "4 + 1" model is a conscious thought appearing in an artificial Working Memory as an emergent property of its complex dynamic structure of memes.

10. The model "4 + 1" can be implemented using a powerful hardware called CBM (Cellular [Automata-based] Brain Machine) being developed by Mikhail Korkin at Genobyte Inc., Boulder, Colorado. The current version of CBM supports almost 75,000,000 neurons organized into modules consisting of up to 1152 neurons. The modules can be evolved using so called CoDi technique I co-develop in the framework of the CAM-Brain Project coordinated at Advanced Telecommunications Research (ATR), Kyoto, Japan. Therefore, regardless its explanational and anticipational powers, the "4 + 1" model is potentially useful as a cognitive structure of human-like robots.

References

1. Andersen SA & Klatzky RL (1987) Traits and social stereotypes: Levels of categorization in person perception, *Journal of Personality and Social Psychology, 53*, 235-246.
2. Anderson JR (1993) *Rules of the Mind*, Hillsdale, NJ: Lawrence Erlbaum.
3. Anderson JR & Bower GH (1973) *Human associative memory*, Washington DC: Winston & Sons.
4. Anderson JR & Lebiere C (1998) *The Atomic Components of Thought*, Mahwah, NJ: Lawrence Erlbaum.
5. Anderson NH (1981) *Foundations of information integration theory*, New York: Academic Press.
6. Asch SE (1946) Forming impressions of personality, *Journal of Abnormal and Social Psychology, 41*, 258-290.
7. Asch SE & Zukier H (1984) Thinking about persons, *Journal of Personality and Social Psychology, 46*, 1230-1240.
8. Brodie R (1996) *Virus of the Mind: The New Science of the Meme*, Seattle: Integral Press.
9. Buller A (1990) A Sub-Neural Network for a Highly Distributed Processing, *8th International Congress of Cybernetics and Systems, June 11-15, 1990, New York City, Conference Proceedings*, The NJIT Press, 185-192.
10. Buller A, Nowak A & Shimohara K (2000) Working Memory and the Buterfly Effect, *XXVII Intenational Congress of Psychology, 23-28 July 2000, Stockholm, Sweden* (accepted for presentation).
11. Buller A & Shimohara K (1999) Decision Making as a Debate in the Society of Memes in a Neural Working Memory, *The Journal of 3D Forum, 13 (3)*, 77-82.
12. Buller A & Shimohara K (2000) Does the 'Butterfly Effect' Take Place in Human Working Memory? *The Fifth International Symposium on Artificial Life and Robotics (AROB 5th '00), January 26-28, 2000, Oita, Japan*, 204-207.
13. Calvin WH (1996) *The Cerebral Code: Thinking a Thought in the Mosaics of the Mind*, Cambridge, Mass: A Bradford Book/The MIT Press.
14. Cantor N & Kihlstrom JF (1987) *Personality and social intelligence*, Englewood Cliffs, NJ: Prentice-Hall.
15. Cantor N & Mischel W (1979) Prototypes in person perception, In: Berkowitz L, Cantor N & Mischel W (Eds.) *Advances in experimental social psychology, Vol. 12.*, New York: Academic Press, 3-53.
16. Churchland PS & Sejnowski TJ (1992) *The Computational Brain*, Cambridge, Mass.: A Bradford Book/The MIT Press.
17. Festinger L (1957) *A theory of cognitive dissonance*, Evanston, IL: Row, Peterson.
18. Fiske ST & Taylor SE (1991) *Social cognition* (2nd ed.), New Yorm: McGraw-Hill.

19. Ford K, Glymour C & Hayes PJ (1995) *Android Epistemology*, Menlo Park: AAAI Press/MIT Press.

20. Gazzaniga MS (Ed.) *The Cognitive Neurosciences*, Cambridge MA: The MIT Press.

21. Gilbert D (1999) Social Cognition, In: Wilson & Keil (1999), 777-778.

22. Haken H (1996) *Principles of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior and Cognition*, Springer: Berlin.

23. Heider F (1958) *The psychology of interpersonal relations*, New York: J.Wiley.

24. Hobson JA (1988) *The dreaming brain*, New York: Basic Books.

25. Hovland C, Janis I & Kelley HH (1953) *Communication and persuasion*, New Haven CT: Yale University Press.

26. James W (1890) *The Principles of Psychology, Vol. Two*, Reprint (1950), New York: Dover.

27. John OP, Hampson SE & Goldberg LR (1991) The basic level of personality-trait hierarchies: Studies of trait use and accessibility in different contexts, *Journal of Personality and Social Psychology, 45*, 20-31.

28. Kelso JAS (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior*, Cambridge, Mass, London: A Bradford Book/The MIT Press.

29. Kim MP & Rosenberg S (1980) Comparison of two structural models of implicit personality theory, *Journal of Personality and Social Psychology, 38*, 375-389.

30. Kunda Z (1999) *Social Cognition. Making Sense of People*, Cambridge, Mass.: A Bradford Book/The MIT Press.

31. Kunda Z & Thagard P (1996) Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory, *Psychological Review, 103*, 284-308.

32. Lewin K (1935) *A dynamic theory of personality*, New York: McGraw-Hill.

33. Matusewicz A (1999) "Model pamięci semantycznej dla autonomicznego agenta", praca magisterska napisana pod kier. A.Bullera, Wydz. ETI, Politechnika Gdańska.

34. Miyamoto Musashi (1645/1982) *A Book of Five Rings*, Woodstock: The Overlook Press.

35. Minsky M (1987) *The Society of Mind*, New York: Simon & Schuster.

36. Nowak A, Lewenstein M & Szamrej J (1993) Bąble modelem przemian społecznych, *Świat Nauki, 12*, 16- 25.

37. Nowak A, Szamrej A & Latané B (1990) From private attitude to public opinion: A dynamic theory of social impact, *Psychological Review, 97*, 362-376.

38. Nowak A & Vallacher RA (1998a) *Dynamical Social Psychology*, Guilford Press.

39. Nowak A & Vallacher R (1998b) Toward Computational Social Psychology: Cellular Automata and Neural Network Models of Interpersonal Dynamics, In: Read & Miller, 277-311.

40. Nowak A, Vallacher RA & Lewnstein M (1994) Toward a Dynamical Social Psychology, In: Vallacher & Nowak (1994a), 279-293.

41. Ostrom TM, Skowronski JJ & Nowak A (1994) The Cognitive Foundation of Attitudes: It's a Wonderful construct, In: Devine PG, Hamilton DL & Ostrom TM (Eds.) *Social cognition: Impact on Social Psychology*, San Diego: Academic Press, 195-258.

42. Plotkin H (1993) *Darwin Machines and the nature of Knowledge*, Cambridge, Mass: Harvard University Press.

43. Posner MI & Keele SW (1968) On the genesis of abstract ideas, *Journal of Experimental Psychology, 77*, 353-363.

44. Read SJ & SJ & Marcus-Newhall A (1993), Explanatory coherence in social explanations: A parallel distributed processing account, *Journal of Personality and Social Psychology, 65*, 429-447.

45. Read SJ & Miller LC (Eds.) (1998) *Connectionist Models of Social Reasoning and Social Behavior*, Mahwah, NJ: Lawrence Erlbaum.

46. Read SJ & Miller LC (1998) On the Dynamics Construction of Meaning: An Interactive Activation and Competition Model of Social Perception, In: Read & Miller (Eds.), 27-68.

47. Rosch E (1978) Principles of categorization, In: Rosch E & Lloyd BB, 27-48.

48. Rosch E & Lloyd BB (1978) *Cognition and categorization*, Hillsdale, NJ: Erlbaum.

49. Rumelhart DE (1980) Schemata: The building blocks of cognition, In: Spiro RJ, Bruce BC & Brewer WF (Eds.), *Theoretical issues in reading comprehension*, Hillsdale NJ: Erlbaum, 249-291.

50. Rumelhart DE, McClelland JL & PDP Research Group (Eds.) (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol.1. Foundations*, The MIT Press: Cambridge, Mass.

51. Schank R & Abelson R (1977) *Scripts, goals, and understanding*, Hillsdale NJ: Erlbaum.

52. Shultz TR & Lepper MR (1996) Cognitive dissonance reduction as constraint satisfaction process, *Psychological Review, 103*, 219-240.

53. Simon H (1995) Machine as Mind, In: Ford et al. (Eds.), 23-40.

54. Smith ER (1996) What connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70*, 893-912.

55. Spellman BA & Holyoak KJ (1992) If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles, *Journal of Personality and Social Psychology, 62*, 913-933.

56. Taylor SE & Crooker J (1981) Schematic bases of social information processing, In: Higgins ET, Herman CP & Zanna MP (Eds.) *Social Cognition: The Ontario Symposium, vol.1*, Hillsdale NJ: Erlbaum, 89-134.

57. Thagard P (1989) Explanatory coherence, *Behavioral and Brain Sciences, 12*, 435-467.

58. Thagard P & Kunda Z (1998) Making sense of people: Coherence mechanisms, In: Read & Miller (1998a), 3-26.

59. Trope Y (1986) Identification and inferential process in dispositional attribution, *Psychological Review, 93*, 239-257.

60. Trope Y & Liberman A (1993) The use of trait conceptions to identify other people's behavior and to draw inferences about their personalities, *Personality and Social Psychology Bulletin, 19*, 553-562.

61. Tulving E (1995) Organization of Memory: Quo Vadis? In: Gazzaniga, 839-847.

62. Vallacher RR & Nowak A (1992) [Temporal patterns in the dynamics of social judgment], Unpublished research data.

63. Vallacher RA & Nowak A (Eds.) (1994a) *Dynamical Systems in Social psychology*, San Diego: Academic Press.

64. Vallacher RA & Nowak A (1994b) The Stream of Social Judgment, In: Vallacher & Nowak (1994a), 251-277.

65. Vallacher RA & Nowak A (1997) The emergence of dynamical social psychology, *Psychological Inquiry, 8* (2), 73-99.

66. Wegner DM & Vallacher RR (1977) *Implicit psychology*, New York: Oxford University Press.

67. Wilson RA & Keil FC (eds.) (1999) The MIT Encyclopedia of the Cognitive Sciences, Cambridge, Mass.: A Bradford Book/The MIT Press.

68. Zubek JP (Ed.) (1969) *Sensory deprivation: Fifteen years of research*, New York: Appleton-Century-Crofts.