Research Proposal for a “Geconcerteerde Onderzoeksactie”

Modelling the Emergence and Evolution of Distributed Cognition

Contents

Summary

This research project is proposed by a multidisciplinary team led by F. Heylighen and F. Van Overwalle. Its members have expertise in cognitive science, psychology, AI, philosophy, economics, sociology and linguistics, and advanced research experience in connectionist simulation, complex systems, self-organization, and group experiments.

The project aims to develop an integrated theory of the emergence of distributed cognition. Distributed cognition is seen as the confluence of collective intelligence, and "situatedness" or the extension of cognitive processes into the physical environment. It concerns the information processing and learning that occurs on the social level, by the propagation of information from agent to agent across media. The theory we wish to develop would have a wide range of social and technological applications, including: better understanding of socio-economic development and diffusion of information, control of cognitive biases and social prejudices, knowledge management and organizational learning, and the development of an intelligent, "semantic" web.

Our approach is based on five working hypotheses inspired by earlier research: 1) groups of agents self-organize to form a coordinated system, 2) the system co-opts external media for transmission of information, 3) the resulting distributed cognitive system can be modelled as a connectionist network, 4) information in the network is propagated selectively, 5) novel knowledge emerges through non-linear, distributed processes. These hypotheses will be elaborated and tested using a combination of theoretical modelling, computer simulation with multi-agent systems and recurrent connectionist networks, and empirical observation, both in controlled laboratory experiments with groups and open-ended observation of "real-world" processes.

1. Presentation of the research team

In the following, the names of the members of the team are printed in **bold** for easy reference. The present proposal is an initiative of Francis **Heylighen** and Frank **Van Overwalle**, supported by their PhD students and research assistants. As such, the research team can be seen as a collaboration between two groups, the *Evolution, Complexity and Cognition* group (ECC), led by **Heylighen**, which is affiliated with the interdisciplinary *Center Leo Apostel* (CLEA), and the *Social Cognition Lab* (SCL), led by **Van Overwalle**, which is affiliated with the *Personality and Social Psychology Department* (PESP).

These two groups have been closely collaborating since 1990 on the dynamics of cognition, with special focus on causal attribution, connectionist learning, and the evolution of shared or collective knowledge. This resulted in several co-authored publications, including [**Van Overwalle** & **Heylighen**, 1991, 1995; **Van Overwalle**, **Heylighen** et al., 1992; **Van Overwalle, Heylighen** & **Heath,** 2004; **Bollen**, **Heylighen** & **Van Rooy**, 1998; **Heylighen**, **Heath** & **Van Overwalle**, 2004].

Heylighen and **Van Overwalle** were moreover co-promotors of two PhD dissertations [**Bollen**, 2001; **Heath**, in preparation], and of several research proposals, including the following :

- Evolutionary Construction of Knowledge Systems (main promotor F. **Heylighen**, funded FWO 1994-1999)
- The Social Construction of Shared Concepts: empirical study and computer simulation of a distributed cognitive process (main promotor F. **Heylighen**, funded FWO 2004-2007)
- Understanding Implicit Learning (main promotor Eric Soetens, submitted GOA 1998)
- Collective Knowledge Development (promotor F. **Heylighen**, submitted FWO 1999)
- Misperception about Groups by Groups (main promotor F. **Van Overwalle**, submitted FWO 2004)
- Mediated Evolution of Social Organisation: a multi-agent simulation (main promotor F. **Heylighen**, submitted FWO 2004)

1.1. Present members of the team

- Prof. Dr. Francis **Heylighen** (promotor, ECC)
- Prof. Dr. Frank **Van Overwalle** (co-promotor, SCL)
- Dr. Bertin **Martens** (researcher European Commission, ECC)
- Tim **Vanhoomissen** (researcher OZR, SCL)
- Carlos **Gershenson** (researcher FWO, ECC)
- Bert **Timmermans** (researcher FWO, SCL)
- Margeret **Heath** (researcher FWO, SCL-ECC)
- Marijke **Van Duynslaeger** (researcher FWO, SCL)
- Andreas **Loengarov** (PhD student, ECC)
- Klaas **Chielens** (PhD student, ECC)

1.2. Former members

- Dr. Johan **Bollen** (associate professor, Computer Science Dept., Old Dominion University (USA), ECC)
- Dr. Dirk **Van Rooy** (lecturer, Psychology Dept., Keele University (UK), SCL)

1.3. Short biographies of the team members

Francis **Heylighen** is a research professor affiliated with the *Department of Philosophy* and the interdisciplinary *Center Leo Apostel* at the *Vrije Universiteit Brussel*. He has worked during most of his career for the *Fund for Scientific Research-Flanders* (FWO), first as research assistant ("aspirant"), then PostDoc, and finally tenured Senior Research Associate ("onderzoeksleider"). He received his MSc in mathematical physics in 1982, and defended his PhD in 1987, on the cognitive processes and structures underlying physical theories [**Heylighen**, 1990]. He then shifted his research to the self-organization and evolution of complex, cognitive systems, which he approaches from a cybernetic perspective.

Dr. **Heylighen** has authored some 90 scientific publications in a variety of disciplines, including a monograph and four edited books. Since 1990 he is an editor of the *Principia Cybernetica Project*, an international organization devoted to the computer-supported, collaborative development of an interdisciplinary knowledge network. He created (and still administers) the project's website [**Heylighen**, Joslyn & Turchin, 2004] in 1993, as one of the first complex, interactive webs in the world. Since 1996 he chairs the *Global Brain Group*, an international discussion forum reflecting on the emerging information society. He is a member of the editorial boards of the *Journal of Memetics,* which he co-founded in 1996, the *Journal of Happiness Studies*, and the journals *Informatica* and *Entropy*.

His work has received a wide and growing international recognition from peers, students and the general public. This is shown by such indicators as the number of references to his work in the combined *Science, Social Science* and *Arts & Humanities Citation Index* (more than 180), on the world-wide web (about 6000 according to www.google.com), in the national and international media (articles about his work have appeared among others in *New Scientist*, *Frankfurter Algemeine Zeitung*, *Die Zeit*, *Le Monde*, the *Washington Post*, and *Knack*), the number of people that have applied to do PhD or PostDoc research under his supervision (several dozen from all around the world), and the invitations he regularly gets to lecture in different countries or to write review articles for leading reference works [e.g. **Heylighen**, 2002; **Heylighen** & Joslyn, 1995, 2001]. He is a Fellow of the *World Academy of Art and Science*, and his biography is listed in *Who's Who in the World* and other international directories.

Frank **Van Overwalle** is a full professor affiliated with the *Department of Psychology* at the *Vrije Universiteit Brussel*. He has worked first as research assistant in the VUB department for new media and computer technology in education, then as PostDoc at the *University of California at Los Angeles* (1988-1989), and finally as PostDoc and tenured professor at the VUB psychology department.

He got his MSc in psychology in 1980, and defended his PhD in 1987 on “Causes of success and failure of freshmen at university: An attributional approach”, for which he received the *Tobie Jonckheere Award* of the *Belgian Royal Academy of Sciences, Letters and Arts*. He continued to work on attribution and social cognition, and then applied his and others’ research to the development of artificial neural network models of social cognition. He has received several grants from his university and the *Fund for Scientific Research-Flanders* in order to test some unique predictions derived from these theoretical proposals. This enabled him to employ several PhD students in his social cognition lab, who generate scientific output either as a PhD or in empirically oriented articles.

Frank **Van Overwalle** has authored some 35 peer-refereed scientific publications, in the domain of social cognition. His recent research focuses on artificial neural network models of various phenomena in the domain of social cognition at large, to demonstrate the common cognitive processes underlying many social findings. The aim is to abolish ad-hoc hypothesis building which is currently very flourishing in social psychology, and to attempt to develop a general cognitive theory encompassing the whole of social psychology, in line with general theories of psychological information processing. This has resulted in a number of publications in top-ranking journals such as *Psychological Review* and *Personality and Social Psychology Review* with an impact score (SSCI) between 3 and 7.

His work is receiving a wide and growing international recognition from peers, as evidenced by some 180 references to his work in the combined *Science, Social Science* and *Arts & Humanities Citation Index*. He is a member of the *Royal Flemish Academy of Art and Science’s* committee of Psychology*,* the *American Psychological Association*, and the executive board of the *Belgian Federation of Psychologists* (BFP). He is a past secretary-general and president of the *Belgian Society of Psychology* (BVP), and is in the editorial board of the *European Journal of Social Psychology* and *Psychologica Belgica*.

Bertin **Martens** is an economist with a MSc (1979) from the *Katholieke Universiteit Leuven*. Since 1989 he works at the *European Commission* in Brussels on project design and evaluation, macro-economic modelling and implementation of structural reform programmes. He has combined his professional career with academic research by working part-time and taking sabbaticals to visit research institutes around the world. As such, he held Visiting Fellow positions at the *University of New South Wales*, the *Max Planck Institute for Research into Economic Systems*, *George Mason University,* and *Stanford University*—where he worked for six months with the Nobel Prize winner Douglas North. He focuses on cognitive science approaches to economic development and institutional change. In May 2004, he defended his PhD thesis [**Martens**, 2004] on the role of distributed knowledge in social and economic evolution, with F. **Heylighen** and M. Despontin as promotors.

Tim **Vanhoomissen** got his MSc (2000) in Experimental Psychology from the *Katholieke Universiteit Leuven*. The goal of his PhD project, under the supervision of **Van Overwalle**, is to develop and test a connectionist model that integrates the important findings of two research fields: perception of groups, and perception of individuals. Using a recurrent model, he has managed to simulate well-known observations from this field, including group-accentuation, illusory correlation in groups, in-group-projection and self-anchoring. The new predictions suggested by this simulation were largely supported by the experiments he undertook to test the model. These results will be presented in his PhD thesis in 2004.

Carlos **Gershenson** is a computer scientist with a BEng (2001) from the *Fundación A. Rosenblueth* in México, and a MSc (2002) from the *School of Cognitive and Computer Sciences* at the *University of Sussex*. He is making a PhD on the design and control of self-organizing systems under the supervision of **Heylighen**, after having been selected from dozens of applicants. His research interests include distributed cognition, philosophy of mind, complex systems, artificial societies and computer simulation. He is a contributing editor to *Complexity Digest*. At the age of 25, he already has published over 20 scientific papers in international proceedings and journals.

Bert **Timmermans** is a psychologist with a MSc (1998) from the *Vrije Universiteit Brussel* and an additional MSc in Cognitive Sciences (1999) from the *Université Libre de Bruxelles*. He is making a PhD under the supervision of **Van Overwalle** on the way summary information is represented and processed in social judgments, and how this can be modelled by a connectionist network. Other fields of interest are implicit learning, neural networks, consciousness, self-consciousness and personality, and artificial intelligence.

Margeret **Heath** is a psychologist with a BA (1989) from the *University of Witwatersrand* in South Africa. She has been doing research as a visiting scholar at different institutes around the world, including the *University of Ottawa*, *International Institute for Management Development* (Switzerland), *St. Gallen University*, the *Santa Fe Institute*, and *George Mason University*. She moreover has many years of experience in business, mostly as a management consultant specialized in facilitating collaboration and analysing and re-designing organizations. She is presently preparing a PhD on the possibility of radical novelty in emerging cognitive systems, under the joint supervision of **Van Overwalle** and **Heylighen**. Her interests include philosophy of mind and imagination, cybernetics, distributed cognition, ethnographic methodologies, and collaborative inventiveness.

Marijke **Van Duynslaeger** studied Clinical Psychology at the *Vrije Universiteit Brussel*. She obtained her MSc in 2002 and an additional MSc in Cognitive Science from the *Université Libre de Bruxelles* in 2003. She is making a PhD under the supervision of **Van Overwalle**, on whether and in what contexts observers spontaneously infer the overt or hidden motives of a person when given information about that person's actions. This research project is funded by the FWO. Her other research interests include attitude formation and persuasive communication.

Andreas **Loengarov** is a sociologist with a MSc (2003) from the *Vrije Universiteit Brussel*. His Master's thesis, presenting a multilevel analysis of the evolution of social and cultural systems, was recognized as outstanding (19/20) by a unanimous jury. He is presently finishing an additional MSc in computer simulation of social systems at the *School of Cognitive and Computer Sciences* of the *University of Sussex*, as a preparation for a PhD under the supervision of **Heylighen**, in which he wants to simulate the role of cultural markers in the evolution of social organization. He has applied to several places for funding.

Klaas **Chielens** is a linguist with a MA (2003) in Germanic philology from the *Vrije Universiteit Brussel*. His Master's thesis [**Chielens**, 2002] made an empirical investigation of selection criteria for the spread of information, and he plans to start working towards a PhD under the supervision of **Heylighen** on the same subject. He has applied to several places for funding.

2. Activities and achievements of the research team

2.1. Previous research

The Evolution, Complexity and Cognition group has been focusing on the self-organization [**Heylighen**, 1988; 2002; **Heylighen** & **Gershenson**, 2003] and evolution [**Heylighen**, **Bollen** & Riegler, 1999] of complex, cognitive systems, such as organisms, groups, societies and computer systems, from a transdisciplinary perspective inspired by systems theory and cybernetics [**Heylighen** & Joslyn, 1995, 2001]. Most of their research is theoretical, aimed at formulating fundamental principles [**Heylighen**, 1992] and integrating conceptual frameworks [**Heylighen**, 2000] to explain the emergence of intelligent organization in such systems. However, this work has also led to concrete technological applications in the design of a self-organizing, "learning" web, that assimilates the implicit knowledge of its users [**Bollen**, 2001; **Bollen** & **Heylighen**, 1998; **Heylighen** & **Bollen**, 2002], and the representation of knowledge through "bootstrapping" semantic and associative networks [**Heylighen**, 2001a, 2001b]. A related strand of work, on the selection criteria that determine which knowledge is transmitted in a large group [**Heylighen**, 1993; 1997; 1998] has received partial empirical confirmation from the statistical analysis of linguistic data [**Heylighen** & Dewaele, 2002; **Chielens**, 2003]. Different models of cognition and learning were also investigated by means of multi-agent computer simulations [**Gershenson**, 2002, 2003, 2004].

The Social Cognition Lab has worked mainly on causal attribution [**Van Overwalle** & **Heylighen**, 1995; **Van Overwalle**, **Heylighen**, Casaer, & Daniëls, 1992; **Van Overwalle** & **Timmermans**, under revision; submitted; **Van Overwalle**, 1989; 1997a, b; 1998], implicit and spontaneous learning and inferences [**Timmermans** & Cleeremans, 2000; **Van Overwalle**, 2004; **Van Overwalle** & **Timmermans**, 2001; **Van Overwalle**, Drenth & Marsman,1999], connectionist modeling of attribution phenomena [**Van Overwalle**,1998, 2003, under revision; **Van Overwalle** & **Van Rooy**, 1998; 2001a, b; **Van Overwalle** & **Timmermans**, 2001] as well as connectionist modeling of social psychology at large. The latter have led to a series of publications on connectionist models, including one publication on group impression formation and biases in Psychological

Review [**Van Rooy**, **Van Overwalle**, **Vanhoomissen**, Labiouse & French, 2003], two publications on person impression formation and cognitive dissonance in Personality and Social Psychology Review [**Van Overwalle** & Jordens, 2002; **Van Overwalle** & Labiouse, 2004] and forthcoming publications on attitude formation [Jordens & **Van Overwalle**, 2004; **Van Overwalle** & Siebler, submitted]. There is recent empirical work supporting some unique predictions of the connectionist approach on group processes and biases [**Vanhoomissen**, De Haan & **Van Overwalle**, submitted] and on attitude formation [Jordens & **Van Overwalle**, 2001; submitted].

2.2. Teaching

Frank **Van Overwalle** teaches three introductory and advanced courses on Social Psychology (with emphasis on social cognition), and one on Group Dynamics. These courses are followed by hundreds of students from different social sciences and humanities. Francis **Heylighen**, as a research professor, only teaches a single course on Complexity and Evolution for some 20 students in philosophy and ethics.

Both have been active in the formation of PhD students, from their own and other departments, by organizing and chairing series of seminars and discussions: the "Foundations Lectures" (1996-2002, **Heylighen**), "CLEA/ECC seminars" (2002-2004, **Heylighen**) and "Boterhammen in de faculteit" (2002-2004, **Van Overwalle**).

2.3. PhD's delivered

Several researchers have prepared and defended their doctorate within the research team, under the (individual or joint) supervision of **Van Overwalle** and **Heylighen**:

- Dirk **Van Rooy** [2000]
- Johan **Bollen** [2001]
- Bertin **Martens** [2004]
- Tim **Vanhoomissen** (defense scheduled Sept. 2004)

The other members of the team are expected to defend their PhD within the next few years.

2.4. Organization of conferences

Both **Heylighen** and **Van Overwalle,** with their collaborators, have organized and chaired several international conferences and workshops on topics related to distributed cognition:

- International Symposium and Workshop on "Self-steering and Cognition in Complex Systems" (VUB, May 20-23, 1987). Proceedings: [**Heylighen**, Rosseel & Demeyere, 1990]
- Summer School on "Self-organization of Cognitive Systems" (Rijksuniversiteit Groningen, Netherlands, August 1988)
- 1st Workshop of the Principia Cybernetica Project: computer-supported cooperative development of an evolutionary-systemic philosophy (VUB, Belgium, July 2-5, 1991)
- Symposium "the Principia Cybernetica Project", as part of the 13th Intern. Congress on Cybernetics (Namur, Belgium, August 1992)

- Symposium “Cybernetic Principles of Knowledge Development”, as part of the 12th European Meeting on Cybernetics and Systems Research, (Vienna, Austria, April 1994)
- Symposium “The Evolution of Complexity,” as part of the international congress “Einstein meets Magritte” (VUB, Belgium, June 1995). Proceedings: [**Heylighen**, **Bollen** & Riegler, 1999]
- 1st Symposium on “Memetics”, as part of the 15th Intern. Congress on Cybernetics (Namur, Belgium, August 1998)
- International Workshop “Classic and Connectionist Approaches to Causal Inference and Social Judgment” (Aix-en-Provence, France, 1999)
- International Workshop “From Intelligent Networks to the Global Brain” (VUB, Belgium, July 3-5, 2001) Proceedings: [**Heylighen** & **Heath**, 2004]
- Workshops on “Social psychology in Belgium” (2002 and 2003).
- One-day International Workshop on “Trends in Distributed Cognition: towards a formulation of a research agenda” (VUB, July 6, 2002)
- Int. Small Group Meeting on “Social Connectionism”, (16-19 June 2004, Genval, Belgium)

2.5. Contacts and collaborations

Francis **Heylighen** and his students actively take part in several international networks related to collective knowledge development and information transmission: The *Principia Cybernetica Project* develops and manages a knowledge web (administered by **Heylighen**) that contains over 2000 documents, including many papers and complete electronic books, which are consulted some 35 000 times a day by people around the world. The *Global Brain Group*, co-founded and chaired by **Heylighen**, groups most of the important researchers in its domain (the emergence of computer-supported, collective intelligence at a world scale), including V. Turchin, B. Goertzel, J. de Rosnay, G. Stock and C. Joslyn. The group organized the first conference on the domain. **Heylighen** administers its electronic mailing list which is used by some 100 selected contributors to discuss advanced issues. **Heylighen** is also involved as founding editorial board member in the *Journal of Memetics: Evolutionary Models of Information Transmission*, where most researchers in the domain publish, and in the *Journal of Collective Intelligence*, which is still in the preparation stage.

His group has been closely collaborating for many years with the *Distributed Knowledge Systems and Modelling* team, led by C. Joslyn at *Los Alamos National Laboratory*, producing several joint publications [e.g. **Heylighen** & Joslyn, 1993, 1995, 2001; Rocha & **Bollen**, 2000]. They also have kept in contact for many years with B. Edmonds in the *Center for Policy Modelling* (Manchester Metropolitan University) and S. Umpleby, director of the *Center for Social and Organizational Learning*, *George Washington University*.

At the international level, Frank **Van Overwalle** collaborates with renowned researchers in the area of connectionist modeling of social phenomena, including Eliot Smith (*Purdue University*, USA), Stephen Read (*USC, Los Angeles)*, Yoshi Kashima (*University* *of Melbourne*, Australia), and Fred Vallée-Tourangeau (*University of Hatfield*, U.K.). He is also a member of a research community of the FWO on "Acquisition and representation of evaluative judgments and emotion". There is also intense collaboration and joint publications with well-known connectionist researchers in other domains of psychology in Belgium, such as at the *Université Libre de Bruxelles* (Axel Cleeremans) and *Université de Liège* (Robert French; Christophe Labiouse).

Locally, within the *Vrije Universiteit Brussel*, our research team maintains and plans to further develop a variety of interdisciplinary contacts, including N. Gontier, E. Myin and J-P. Van Bendegem at the Center for Logic and Philosophy of Science (CLWF) on the evolution of language and the extended mind, E. Verstraeten and E. Soetens of the Cognitive and Physiological Psychology group (COPS) on brain physiology and implicit learning, G. Geeraerts and K. Laforce of the Political Science Department (POLI) on complex systems models of social interaction, T. Belpaeme of the AI-lab on computer simulations of cognitive and language evolution, K. Tuyls and A. Nowé of the Computational Modelling Lab (COMO) on multi-agent systems, and G. Vancronenburgh and N. Deschacht of the Economics Department on evolutionary and systems dynamics models of social and economic interaction.

At our sister university, the *Université Libre de Bruxelles*, we plan to stay in touch with T. Lenaerts of the AI-lab (IRIDIA) on the evolution of cooperation, A. Cleeremans of the Cognitive Science Research Unit on connectionist models of cognition, the group around J-L. Deneubourg at the Center for Non-Linearity and Complex Systems on insect models of collective intelligence, and O. Klein at the Social Psychology Department on communication and maintenance of stereotypes in groups.

3. Introduction to the research theme

Cognition can be defined as the collection and processing of information in order to support decision-making and problem-solving by an agent. The cognitive agent uses its *knowledge* to interpret incoming data or stimuli, derive inferences from it, and select actions appropriate to the thus perceived situation and to its internal preferences. This knowledge is in general the result of previous *learning*, i.e. adapting the internal structure responsible for processing the information so as to maximize the quality of the inferred predictions and selected actions, while taking into account the feedback from the environment. From this "cybernetic" perspective [**Heylighen** & Joslyn, 2001; **Van Overwalle** & **Van Rooy**, 1998; **Van Overwalle** 1998; **Van Overwalle** & Labiouse, 2003], knowledge is not a discrete collection of beliefs, propositions or procedures, but a continuously evolving relation between perceptions, inferences and actions, which allows the agent to anticipate and adapt to changes in its environment [**Heylighen**, 1990].

The study of cognition—*cognitive science*—is in essence multidisciplinary, integrating insights from approaches such as psychology, philosophy, artificial intelligence (AI), linguistics, anthropology, and neurophysiology. To this list of sciences of the *mind*, we now also must add the disciplines that study *society*. Indeed, an increasing number of approaches are proposing that cognition is not limited to the mind of an individual agent, but involves interactions with other minds.

Sociologists have long noted that most of our knowledge is the result of a social construction rather than of individual observation [e.g. Berger & Luckman, 1967]. Philosophers have brought the matter to research for urgent consideration in theories of mind [e.g. Searle, 1995]. The nascent science of memetics [Aunger, 2001; **Heylighen**, 1998], inspired by evolutionary theory and culture studies, investigates the spread of knowledge from the point of view of the idea or *meme* being communicated between individuals rather than the individual that is doing the communication. Economists too have started to study the role of knowledge in innovation, diffusion of new products and technologies, the organization of the market, and overall social and economic development [**Martens**, 1998, 2004]. Management theorists emphasise knowledge management and learning as an organisational phenomenon rather than as an individual process. Effective organisational learning is deemed to be the difference between an enterprise that flourishes and one that fails [Senge, 1990]. Social psychologists have started to do laboratory experiments to study cognition at the group level [e.g. Brauer et al., 2001; Klein et al., 2003; **Van Rooy, Van Overwalle** et al., 2004]. Biologists, aided by computer scientists, have built models that demonstrate how collectives of simple agents, such as ant colonies, bee hives, or flocks of birds, can process complex information more effectively than single agents facing the same tasks [Bonabeau et al., 1999]. Building on the tradition of distributed artificial intelligence, the subject of collective cognition is now even being investigated mathematically [Crutchfield et al. 2002].

These different approaches provide a new focus for the understanding of cognition that might be summarized as *collective intelligence* [Levy, 1997; **Heylighen**, 1999], i.e. the cognitive processes and structures that emerge at the social level. But at the same time the investigation of cognition has expanded in another direction: that of the physical environment.

The failure of traditional, "symbol-processing" AI to come up with workable models of intelligence has pointed to the necessity for *situatedness, embodiment* or *enaction* [Steels & Brooks, 1995; Clark, 1997], i.e. the observation that cognition or mind cannot exist in a mere abstract realm of ideas (the "brain-in-a-vat"), but must be part of an interaction loop, via perception and action, with a concrete environment [cf. **Heylighen** & Joslyn, 2001]. This has led to a flurry of interest in autonomous robots which forego complex representations and symbol manipulations by using the environment as its own best model [Steels & Brooks, 1995].

The environment supports cognition not just *passively*—by merely representing itself, but *actively*—by registering and storing agent activities for future use, and thus functioning like an external memory [Kirsh, 1996; Kirsh & Maglio, 1994; Clark, 1997]. Examples abound, from the laying of pheromone trails by ants and the use of branches to mark foraging places by wood mice to the notebooks we use to record our thoughts. Physical objects can further be used to collect and process information, as illustrated by telescopes and computers. This use of external phenomena as "epistemic structures" [Kirsh & Maglio, 1994] that support internal information processing leads to a view of cognition expanding outside the brain: the *extended mind* [Clark & Chalmers, 1998]. This is an active form of the philosophy of *externalism*, according to which external phenomena take part in mental content.

The "offloading" of information onto the environment makes this information potentially available for other agents, thus providing a medium by which information sharing, communication, and coordination can occur. This basic mechanism, known as *stigmergy*, underlies many examples of collective intelligence [Bonabeau et al., 1999; **Heylighen**, 1999; Susi & Ziemke, 2001], such as the trail laying of ants and the mound building of termites. More generally, any form of information exchange between agents requires the use of external media, such as sound waves, light, or electrical signals. Thus, the two perspectives of collective intelligence and situatedness necessarily tie in with each other.

They can be integrated under the heading of *distributed cognition* [Hutchins, 1995]: in many systems, information processing and knowledge generation takes place across agents and physical media, forming an integrated cognitive system whose processes and structures are distributed over all its components. Functional relationships between the components create a computational system that solves real world problems. The central idea is that the processing occurs through what Hutchins [1995] calls *the propagation of representational states across representational media*. Hutchins, and his collaborators at UCSD and Indiana (Kirsh, Hollan, Maglio et al) have begun to develop highly refined ethnographic research methodologies in order to map what they call Wild or Raw cognition, i.e. information processing as it happens in the real world rather than in a laboratory set-up or computer simulation. The paradigmatic example investigated in detail through this methodology is the navigation of a large ship, which requires the activity of several people coordinated by means of instruments, ship navigation manuals, communication channels, and the necessary enacted/ situated deviation from guidelines and formal process [Hutchins, 1995].

4. Aim: towards an integrated theory of distributed cognition

In spite of its promises, the distributed cognition approach as yet offers little more than a heterogeneous collection of ideas, observation techniques, preliminary simulations and case studies. It lacks a coherent theoretical framework that would integrate the various concepts and observations, and provide a solid foundation for building detailed models of concrete systems and processes [**Heylighen**, **Heath** & **Van Overwalle**, 2004; Susi & Ziemke, 2001]. The present proposal aims to develop such an integrated theory, supported by observations, experiments and detailed computer simulations.

For us, understanding distributed cognition at the deepest level requires understanding how it *originates*. The analysis of existing distributed processes, such as ship navigation, is not sufficient, because the underlying systems tend to be constrained and specialized, while their often convoluted way of functioning is typically rigidly set as the result of a series of historical accidents. A more general understanding, not only of the "how?" but also the "what?" and the "why?", may be found by analysing how distributed cognition *emerges* and *evolves* step by step in a system that initially does not have any cognitive powers. We wish to focus on the *creation*—and not merely the propagation—of knowledge and information in these systems.

Our basic research questions can be formulated as follows:

- How do initially independent agents through interaction (using external media) come to form a distributed cognitive system?
- What kind of coordination between their different information processing activities emerges?
- What knowledge is novel or emergent in this system, i.e. knowledge that did not already exist in the mind of an individual agent?
- In what way is this emergent cognition better or worse than the initial, individual cognition?
- More specifically, which information is lost or filtered out during the process?
- Which features influence the efficiency of the process? For example, in how far do the resulting cognitive capabilities depend on the number of agents, the diversity in experience between agents, or the presence or absence of different types of media?

5. Social relevance and potential applications

A theory of distributed cognition as we envisage it here would offer a wealth of potential applications, with particular relevance to society at large. To start with, understanding how knowledge and information are distributed throughout social systems would help us to foster the economic and social development that new knowledge and better coordination engenders [**Martens**, 1998; 2004]. In particular, such a theory should tell us how important new ideas can diffuse most efficiently, and conversely how the spread of false rumours, superstitions and "information parasites" might be curtailed [**Heylighen**, 1999]. More generally, it may help us to control for the cognitive biases and social prejudices whose ubiquity psychologists have amply demonstrated [Brauer et al., 2001; Klein et al., 2003; **Van Rooy**, **Van Overwalle** et al., 2004].

On a smaller scale, a theory of distributed cognition has immediate applications in business, government, and other organizations. It would help them to promote innovation and avoid the pitfalls of collective decision-making, such as *groupthink* [Janis, 1972], which stifle creativity. It would support organizations not only in generating new knowledge but in efficiently maintaining, applying and managing the knowledge that is already there. More fundamentally, it would provides us with concrete guidelines to design more effective organizations, where roles and functions are clearly specified, and where information is processed in a coordinated way, with a minimum of loss, distortion, misunderstanding or confusion. In sum, it would foster the collective intelligence of the organization, while minimizing the inherent tendency of groups towards "collective stupidity".

Technological applications abound as well. A crucial application of the proposed model of distributed cognition would be the compilation by committees of experts of formal "ontologies" [Staab & Studer, 2003], i.e. the systems of categories necessary for the *semantic web* [Berners-Lee et al., 2001]. This knowledge architecture for the future Internet will allow users to get concrete answers to specific questions, while enabling various services to automatically coordinate. But this requires efficient and consensual schemes to represent knowledge that is generated and managed in a distributed manner. More generally, a lot of research is going on in distributed AI to develop efficient coordination schemes to let software agents collaborate. One of the more immediate application domains is *ambient intelligenc*e [ISTAG, 2003]. This refers to the vision of everyday artefacts and devices such as mobile phones, coffee machines and fridges exchanging information and coordinating with each other so as to provide the best possible service to the user, without needing any programming or prompting—thus effectively extending the user's mind into his or her physical environment [**Gershenson** & **Heylighen**, 2004].

Integrating the ambient intelligence of devices, the collective intelligence of organizations and society, and the global communication and coordination medium that is the future Internet leads us to a vision of a *global brain* [**Heylighen** & **Bollen,** 1996; **Heylighen**, 1999; **Heylighen** & **Heath**, 2004], i.e. the emerging intelligent network formed by the people of this planet together with the knowledge and communication technologies that connect them together.

6. Working hypotheses

Inspired by our earlier research, we propose five fundamental *working hypotheses*, to function as starting points or postulates out of which we will try to build a general model of distributed cognition.

6.1. groups of agents self-organize

Consider a group of initially autonomous actors, actants or *agents*, where an agent can be human, animal, social or artificial. Agents by definition perform *actions*. Through their shared environment the action of the one will in general affect the other. Therefore, agents in proximity are likely to *interact*, meaning that the changes of state of the one causally affect the changes of state of the other. These causal dependencies imply that the agents collectively form a *dynamical system*, evolving under the impulse of individual actions, their indirect effects as they are propagated to other agents, and changes in the environment. This system will typically be non-linear, since causal influences normally propagate in cycles, forming a complex of feedback loops. Moreover, a dynamical system has computational structure and is therefore in principle able to process information and generate patterns [Crutchfield, 1998].

While such a complex system is inherently very difficult to model, control or predict, all dynamical systems tend to self-organize [Ashby, 1962; **Heylighen** & Joslyn, 2001; **Heylighen**, 2003; **Heylighen** & **Gershenson**, 2003], i.e. evolve to a relatively stable configuration of states (an *attractor* of the dynamics). We can say that the agents in this configuration have mutually adapted [Ashby, 1962], limiting their interactions to those that allow this collective configuration to endure. There is further an on-going selective pressure to make these interactions more synergetic [Wright, 2000; **Heylighen**, 2004], because a mutually beneficial interaction is preferable to one that is less so. In this view, the self-organization and further evolution of the collective configuration effectively creates a form of social *organization*, in which agents support each other's activities so as to maximize their collective benefit. This configuration can be viewed as a *mediator*, coordinating the agents' actions [**Heylighen**, 2004]. As illustrated by the many simulations of the evolution of cooperation [e.g. Axelrod, 1984; Riolo, Cohen & Axelrod, 2001; Hales & Edmonds, 2003], this allows them to overcome the obstacles of individual selfishness or "free riding" that are exemplified by the Prisoners' Dilemma [Axelrod, 1984; **Heylighen**, 1992; **Heylighen** & Campbell, 1995], .

According to *coordination theory* [Crowston, 2003], we can distinguish the following fundamental dependencies between activities or processes in an organization: 1) two processes can use the same resource (input) and/or contribute to the same task or goal (output); 2) one process can be prerequisite for the next process (output of the first is input of the second). The first case calls for tasks to be performed in parallel and the second case in sequence. Efficient organization means that the right activities are delegated to the right agents at the right time. The parallel distribution of tasks determines the *division of labor* between agents. The sequential distribution determines their *workflow*.

Division of labor reinforces the specialization of agents, allowing each of them to develop an expertise that the others do not have [Gaines, 1994; **Martens**, 2004]. This enables the collective to overcome individual cognitive limitations, accumulating a much larger amount of knowledge than any single agent might. Workflow allows information to be propagated and processed sequentially, so that it can be refined at each stage of the process. Self-organization thus potentially produces emergent cognitive capabilities that do not exist at the individual level.

6.2. the organization co-opts external media for information sharing

Self-organization in this sense can be seen as the more efficient, synergetic use of interactions. Interactions between agents necessarily pass through their shared physical environment. We will call the external phenomena that support these interactions *media*. Certain parts or aspects of the environment lend themselves better to synergetic interaction than others do. For example, a low-bandwidth communication channel that is difficult to control, such as smoke signals, will support less synergetic interactions than a reliable, high-bandwidth one, such as optical cable. Thus, there is a selective pressure for agents to preferentially use the more efficient media, i.e. the ones through which causal influences—and therefore information—are transmitted most accurately and reliably.

Moreover, simply by using them, the agents will change the media, generally adapting them to better suit their purposes. For example, animals or people that regularly travel over an irregular terrain between different target locations (such as food reserves, water holes or dwellings) will by that activity erode paths or trails in the terrain that facilitate further movement. The paths created by certain agents will attract and guide the movements of other agents, thus providing a shared coordination mechanism that lets the agents communicate indirectly. Thus, actions (trajectories of movement) and media (tracks eroded in the terrain) co-evolve, the one adapting to better fit the other. A slightly more advanced version of this mechanism are the trails of pheromones laid by ants to steer other members of their colony to available food sources, thus providing the colony with a *collective mental map* of its surroundings [**Heylighen**, 1999]. Humans, as specialized tool builders, excel in this adaptation of the environment to their needs, and especially in the use of physical signs and symbols, electromagnetic waves, or hardware to store, transmit and process information.

In this way, external media are increasingly assimilated or co-opted into the social organization, shaping it while being shaped by it, and making the organization's functioning ever more dependent on them. As a result, the collective cognitive system is extended into the physical environment and can no longer be separated from it.

6.3. distributed cognitive systems function like connectionist networks

Considering an extended social organization or distributed cognitive system at the most abstract level, we can distinguish *nodes*, i.e. the agents or objects that store or contain information, and *links*, i.e. the media or channels along which information is transmitted between nodes. Links can have variable *strength*, where strength represents the ease, frequency or intensity with which information is transmitted. They represent stabilized causal influences between agents and/or objects, possibly supported by co-opted media.

Every node is characterized by its space of possible states. The actual state at the beginning of a process is propagated in parallel along the different links, and recombined in the receiving nodes. State spaces can in general be factorized into independent variables or degrees of freedom, each of which can take on a continuum of values [**Heylighen**, 2002]. A complex node can thus be functionally decomposed as an array of simple, one-dimensional nodes that only take on a single "intensity" or "activation" value. The resulting network of simple nodes and links appears functionally equivalent to an "artificial neural network", or what we prefer to call a *connectionist network*, where activation spreads from node to node via variable strength links [**Van Overwalle** & Labiouse, 2004; McLeod et al., 1998]. This network is in general *recurrent*, because of the existence of cycles or loops as mentioned earlier.

Connectionist networks have proven to provide very flexible and powerful models of cognitive systems [e.g. McLeod et al., 1998; **Van Overwalle** & Labiouse, 2004; **Timmermans** & Cleeremans, 2000]. Their processing is intrinsically parallel and distributed [Rumelhart & McClelland, 1986]. Because of the inherent redundancy, they are much more robust than sequential architectures, surviving destruction of part of their nodes and links with merely a "graceful" degradation of their performance. These systems are wholly decentralized and self-organizing, eliminating the need for a central executive that deliberatively processes information. Moreover, since activation spreads automatically from the nodes that received the initial stimuli to associated nodes, connectionist networks exhibit emergent phenomena such as pattern completion and generalization. This allows them to fill in lacking data and to infer plausible conclusions on the basis of very limited information.

Most importantly, connectionist networks inherently support learning, by means of the continuous adaptation of the link strengths to the ways in which they are used. Thus, successfully used links become stronger, making it easier for information to be propagated along them, while links that are rarely used or whose use led to erroneous results weaken. In an extended cognitive system we can conceive of at least two mechanisms for such selective reinforcement. On the physical level, commonly used media become more effective, as proposed in the previous hypothesis. But a more flexible mechanism is social adaptation, in which an agent learns from the experience of communicating with another agent. If the other agent reacts appropriately, the first agent will increase its trust in the other's competence and goodwill, and thus becomes more likely to communicate similar information to that agent in the future.

As such, the network's "experience" of use is stored in long-term weight changes of its connections. Thus, the network acquires new knowledge in a distributed manner, i.e. storing it in the pattern of links rather than in the states or memories of individual nodes. An example of such a distributed learning system is the *invisible hand* of the market, which "knows" how to make supply match demand by allocating resources to the agents that appear most competent to satisfy the demand [**Heylighen**, 1997].

6.4. information in the network is propagated selectively

Whether information is transmitted will not only depend on the architecture of the network, but on the content of the information. Memetic analysis and social-psychology observation have suggested different selection criteria that specify which information is preferentially passed on [**Heylighen**, 1993, 1997, 1998]. These include the criteria of:

- *utility* (the information is useful or valuable to the agents)
- *novelty* (the information is not already known)
- *coherence* (the information is consistent with the knowledge that the agents already have)
- *simplicity* (since complex information is difficult to process, less important details tend to be left out)
- *formality* (the less context or background communicating agents share, the more important it is to express the information explicitly)
- *expressivity* (the information is easily expressible in the available media)
- *authority* (the source is recognized as being trustworthy)
- *conformity* or *consensus* (the majority of agents agree on the information)

Several of these criteria have been empirically confirmed through psychological experiments [Lyons & Kashima, 2003] and analysis of linguistic data [**Heylighen** & Dewaele, 2002; **Chielens**, 2003]. They provide a simple set of guidelines to understand the evolution of distributed knowledge through variation and selection [**Heylighen**, 1993, 1998].

A theory of distributed cognition would ideally allow these criteria to be derived from the dynamics of a distributed connectionist network, rather than have them posited to some degree *ad hoc*. A preliminary simulation [**Van Overwalle**, **Heylighen** & **Heath,** 2004] indeed suggests that this can be achieved. For example, the reinforcement of links through the increase of trust builds authority for the sending agents, while telling them which information the receiving agents are likely to already know and agree with, making it less important for them to transmit detailed, explicit reports. Moreover, spread of activation along existing connections will automatically attenuate inconsistent [**Van Overwalle** & Jordens, 2002] or complex signals, while amplifying signals that are confirmed by many different sources (conformity) or that activate in-built rewards or punishments (utility).

Selective propagation and thus filtering out of less relevant or less reliable data already constitutes information processing, as it compresses the data and thus potentially distils the underlying pattern or essence. However, if selectivity is inadequate, this can lead to the loss of important ideas, and the propagation of incorrect information, as exemplified by the flurry of social and cognitive biases that characterizes "groupthink" [**Van Rooy**, **Van Overwalle, Vanhoomissen** et al., 2003]. More extensive modelling and simulation should allow us to identify the central factors through which we can control these dangerous tendencies.

6.5. novel knowledge emerges

On the positive side, groups often are more intelligent than individuals, integrating information from a variety of sources, and thus overcoming individual biases, errors and limitations. In the simplest case, this occurs through a superposition of individual contributions. Because of the law of large numbers, the larger the variety of inputs, the smaller the overall effect of random errors, noise, or lacking data, and the clearer and more complete the resulting collective signal [**Heylighen**, 1999]. This "averaging" of contributions is represented very simply in a connectionist network, by the activation from different inputs being added together and renormalized in the target nodes.

But a recurrent connectionist network, being non-linear and self-organizing, may offer more radical forms of novelty creation, through the emergence of structures that are more than the sum of their parts. Rather than being attenuated by averaging, noise can here play a creative role, triggering switches to a wholly new attractor or configuration at the bifurcation points of the dynamics, thus exemplifying the "order from noise" principle [von Foerster, 1960; **Heylighen**, 2002; **Heylighen** & **Gershenson**, 2003].

The same mechanisms of self-organization that lead to coordination between agents are also likely to lead to coordination and integration of the ideas being communicated between those agents. An idea that is recurrently communicated will undergo a shift in meaning each time it is assimilated by a new agent, who adds its own, unique interpretation and experience to it. Moreover, the need to express it in a specific medium will also affect the shape and content of the message, which will be further constrained by the need to achieve an invariant external reference or "intentionality" for it [Cantwell Smith, 1996]. Like in a game of Chinese whispers [cf. Lyons & Kashima, 2003], by the time the idea comes back to the agent who initiated it, it may have changed beyond recognition. After several rounds of such passing back and forth between a diverse group of agents, the dynamical system formed by these propagations with a twist is likely to have reached an attractor, i.e. an invariant, emergent configuration.

In this way, novel shared concepts may self-organize through communication, providing a basic mechanism for the social construction of knowledge [Berger et al., 1967]. Concrete illustrations of this process can be found in multi-agent simulations of the origin of language where the symbol (external support) co-evolves with the category that it refers to (internal concept with external reference) [e.g. Hutchins & Hazelhurst, 1995; Steels, 1998; Belpaeme, 2001]. These models are based on recursive *language games*, where a move consists of one agents expressing a concept and the receiving agent indicating whether or not it has "understood" what the expression refers to (e.g. by pointing towards a presumed instance of the category), after which the first agent adjusts its category and/or expression. After a sufficient number of interaction rounds between all the agents in the collective, a "consensus" typically emerges about a shared concept and its expression.

Knowledge consists not only of concepts or categories, but of logical and causal connections between these categories. These have the general form [**Heylighen**, 2001a]:

> IF occurrence of category *A* (e.g. *banana* or *lack of preparation*),
> THEN expect occurrence of category *B* (e.g. *yellow* or *failure for exam*).

Such basic connections underlie not only expectation or prediction, but causal attribution or explanation of *B*, given *A* [**Van Overwalle** & **Heylighen**, 1991, 1995; **Van Overwalle**, 2003]. The connections between categories can be learned through the closely related *Hebbian* [e.g. **Heylighen** & **Bollen**, 2002] or *Delta algorithms* [**Van Overwalle**, 1998, 2003; **Van Overwalle** & **Van Rooy**, 1998, 2001a,b]. These connectionist learning rules are simple and general enough to be applicable even when cognition is distributed over different agents and media [e.g. **Heylighen** & **Bollen**, 2002; **Bollen**, 2001; **Van Overwalle**, **Heylighen** & **Heath**, 2004], as argued in 6.3. However, if we moreover take into account the social construction of concepts, we get a view of concepts, symbols, media and the connections between them co-evolving, in a complex, non-linear dynamics. This points us towards a potential "bootstrapping" [**Heylighen**, 2001a] model of how complex and novel distributed cognitive structures, such as languages, scientific theories, world views and institutions, can emerge and evolve.

7. Methodologies for distributed cognition research

The study of distributed cognition is in essence multidisciplinary, and our research therefore will need to integrate methods from very different traditions, including theoretical analysis and model-building, computer simulation and empirical observation.

7.1. Theoretical investigation

The very wide variety of existing models, concepts and observations makes it clear that in order to elaborate our working hypotheses into a full theory we first of all need to focus on the collection and theoretical integration of existing models and observations. This will require an extensive review of the relevant literature in the many related disciplines, and the consultation of a variety of domain experts.

Happily, our team has the required multidisciplinary expertise, its members having degrees in cognitive science, psychology, computer science, sociology, linguistics and economics; advanced research experience in philosophy, cybernetics, connectionism, management and complex systems; and local and international contacts with a range of specialists in the relevant research topics. Moreover, we have extensive experience in interdisciplinary integration [e.g. **Heylighen**, 1992; 1990b], sometimes in the form of connectionist models [e.g., **Van Overwalle**, 1998; **Van Overwalle** & Jordens, 2002; **Van Overwalle** & Labiouse, 2003], and in both traditional (workshops, seminars, ...) and computer-supported forms (mailings lists, web-based discussion forums, ...) of intellectual discussion and collaboration [**Heylighen**, 2000].

A more specific methodology for theoretical research that is increasingly popular among philosophers is the *thought experiment*: imagine a system with such and such characteristics, put in such and such circumstances; what will happen? Different models and approaches will typically make different predictions. Theoretical analysis and inference will then allow us to find out in what respect the models agree or disagree, highlighting their similarities and differences and thus giving us a common basis to integrate them. A well-chosen thought experiment may moreover help us to find out that certain models are incoherent (self-contradictory), inconsistent with known facts, or simply incomplete and ambiguous. This will help us to focus on the issues that need to be investigated further, or complemented by other approaches.

7.2. Computer simulation

A more advanced version of a thought experiment is a computer simulation [**Gershenson**, 2002a]. Here we make the theoretical model sufficiently explicit so that its rules can be programmed. The advantage is that the computer can explore many more possible combinations of initial conditions, and infer many more of their consequences than a theoretician can. Thus, a well-designed simulation platform can provide us with a true *virtual laboratory* [**Gershenson**, González & Negrete, 2000], which we can use to quickly and easily test thousands of variations on a basic model simply by varying the parameter values. Such a virtual laboratory can even be used to compare the predictions of fundamentally different paradigms for modelling cognition, such as dynamical systems, connectionist networks and rule-based systems, by programming agents to behave according to each of the models and then registering in what way their concrete behaviors differ [**Gershenson**, 2003, 2004].

Work in the collective intelligence/distributed AI tradition has typically relied on *multi-agent simulations* (MAS), in which interacting software agents form a kind of "artificial society" (Bonabeau et al., 1998). An alternative simulation paradigm are the connectionist networks, which tend to give more precise, numerical predictions than MAS, but tend to be less effective in providing an intuitive, qualitative understanding of the system that is modelled. Our research team has extensive experience with both types of simulations [e.g. **Gershenson**, 2003; **Van Rooy**, **Van Overwalle**, **Vanhoomissen** et al. 2003], and has recently started to explore an integrated framework where connectionist agents interact through "extended" communicative connections, as proposed in hypothesis 6.3 [**Van Overwalle**, **Heylighen** & **Heath**, 2004].

7.3. Observation

The disadvantage of simulations is that they still are based on a very simplified model of reality, which is wholly dependent on the subjective assumptions of the designer. Therefore, many simulations have been criticized for merely confirming the biases of their creators. Real-life observations of actual social systems, as used in the distributed cognition tradition [Hutchins, 1995], can evade these criticisms, by providing an open-ended source of unanticipated effects and interactions. The disadvantage is that they are very time-consuming and difficult to control so that only a few variations of a basic situation can be investigated.

We therefore wish to combine the benefits of both methodologies, using observation to suggest new hypotheses and simulation to quickly explore the different implications of these hypotheses, so that the most promising ones can become the focus of a new observation. Moreover, the results from the observations can be used to adjust the parameters of the simulation, as we have frequently done with our connectionist simulations of individual and group cognitive processes [e.g. **Van Rooy**, **Van Overwalle**, **Vanhoomissen** et al., 2003; **Van Overwalle, Heylighen & Heath,** 2004]. Conversely, since the simulation can be run with many different rules and initial conditions, this may allow us to find the most interesting cases (e.g. that demarcate different models), which we can then try to replicate empirically.

There are two basic methods of empirical data gathering relevant for distributed cognition research: *experiments*, in which the set-up is explicitly manipulated by the researcher to control for specific variables, and *ethnographic observations* or case-studies, in which the researcher investigates an existing system, trying to interfere as little as possible, while noting down all observed phenomena. The former methodology is most common in psychology, the latter in cognitive anthropology [e.g. Hutchins, 1995] and organizational studies. Our group has experience with both approaches, especially with laboratory experiments [e.g. **Van Overwalle, Heylighen** et al., 1992**,** Jordens & **Van Overwalle**, 2001; **Van Overwalle** & **Van Rooy**, 2001a, b; **Van Overwalle**, Drenth & Marsman, 1999], but also with video recording and content analysis of group problem solving sessions, and the statistical analysis of existing linguistic corpora (e.g. recordings of conversations [**Heylighen** & Dewaele, 2002], or virus hoaxes [**Chielens**, 2003]).

Compared to "field" observations, experiments provide more explicit control over different conditions, so that they allow us to test and compare different models more precisely. However, by creating an artificial, researcher-designed situation, they may ignore real-world, "wild" phenomena [Hutchins, 1995]. As such, experiments can fill the gap between the open-ended but difficult to control field observations and the "closed" computer simulations. We will now propose two experimental paradigms that try to combine the advantages of both approaches in investigating distributed cognition.

7.4. Group Communication Experiments

The more traditional psychological experiments where individual participants are subjected to controlled stimuli (e.g. flashes of light, or reading a text) after which their reactions are registered (e.g. by letting them fill in a questionnaire concerning their experience/ interpretation) appear ill-suited for observing distributed cognitive processes since these essentially occur *between* participants. However, such interaction between subjects is increasingly being studied through group experiments where members of a group work towards a common goal, or observe the same stimuli, after which the individual (private) or group (public) reaction is measured.

Thus, to test hypotheses concerning group interaction, we can focus on minute details of the interaction between participants, extending earlier research on the individual's reaction to private stimuli into the realm of group input. For instance, at the individual level, we might explore to what extent *authority* and *trust* in other group members and the ideas that they express, are psychological phenomena that arise automatically during a conversation or discussion, or whether this is a more consciously controlled event. Alternatively, at the group level, we can explore how controlled information may be shared or become distorted during the communication or discussion in a group.

For this latter type of experiments, social psychologists have used two basic paradigms that reflect alternative ways in which information is propagated between people: *parallel* and *serial* communication. In a parallel communication design, the information is spread from all communicators directly to each participant, like in a group discussion. Thus, the participant has direct access to the observations and impressions of all the people who received the stimuli. The communicators spread information that contains, for instance, consistent and inconsistent behaviors relevant to a target group. Afterwards, the participants provide their own impressions about the target groups. In a serial reproduction design, the communication of information is passed sequentially from person to person, like in rumors and gossip. The first communicator in the chain receives the information, memorizes it, and then communicates this information to the second person in the chain, and so on. Here we can investigate how the information changes as it progresses through the chain, depending on factors such as the background knowledge that the participants have [e.g. Lyons & Kashima, 2003].

7.5. Computer-mediated games

A related experimental paradigm, inspired by MAS, experimental economics, and studies of group dynamics, may provide us with a direct bridge between empirical and simulation methods. Most MAS and economics experiments have the structure of a "game" where agents (people or software agents) interact by making "moves" towards their partners, following certain imposed constraints or rules, while trying to achieve an individual or collective goal (e.g. maximizing their utility).

Usually, these games (e.g. the ubiquitous Prisoners' Dilemma game) are rigidly constrained, leaving the agents very little freedom in choosing what move to make (e.g. either "cooperate" or "defect"). This creates a highly artificial situation whose relevance to real-world phenomena is limited. However, this does not need to be the case, as we can conceive a continuum of game situations, from completely controlled to almost completely free-form and spontaneous. Free-form games (e.g. unconstrained brainstorming sessions) may attract our attention to unanticipated phenomena, while more constrained games allow us to test specific hypotheses and compare different models or parameter values.

Still, even free, open-ended games can give us accurate control over data collection. Suppose we let the participants interact through a computer-supported medium that offers them a specific choice of moves. The computer system registers which moves were made by whom at what moment, providing the experimenter with precise, easily analyzable data. For example, the system may support a group discussion by allowing the participants to submit specific types of contributions: propositions, questions, confirmations, refutations, evaluations, etc. However, the system should also allow completely free-form, unconstrained interactions (e.g. spoken and non-verbal communication) that can be recorded on video for content analysis, so as not to artificially restrict expression.

Such a group discussion does not need to be limited to experimenter-defined topics or formats, but can include real-world activities, such as the scientific discussions that form the basis of the Principia Cybernetica Project [**Heylighen**, 2000]. In this case, the observers do not control the topic, participants or dynamics of the discussion, but merely offer tools to assist the participants in their spontaneous interactions, while using those tools to accurately register what happens.

The advantage of the more constrained computer-mediated games, on the other hand, is that they lend themselves to direct comparison with multi-agent simulations. For the more rigidly defined games (such as the Prisoners' Dilemma) it is easy to run the same game with software agents and human subjects, so that the similarities and differences between simulation and reality can be evaluated numerically.

8. Concrete subprojects

We will now show how these methodologies can be applied to test and elaborate each of the five working hypotheses (sections 6.1 - 6.5) that form the backbone of our proposal. This defines five concrete subprojects within our overall proposal for the development of an integrated theory of distributed cognition.

8.1. groups of agents self-organize

While all dynamical systems will eventually "self-organize" (reach an attractor) by definition [Ashby, 1962], the concrete question we must address is how and under what conditions a group of agents will self-organize, and what kind of cognitive or social structures will emerge from their interactions. Given the complexity of this process, and the many steps that can be expected to be necessary in order to see non-trivial structures emerge, this working hypothesis is best tested through an agent-based computer simulation. However, the rules and assumptions for this simulation will need to be based on a theoretical analysis and review of existing models and observations.

To build the simulation, we plan to start from the KEBA (Knowledge Emerging from Behavior) system that we have developed earlier [**Gershenson**, 2002]. This is a 3D, virtual environment where agents interact with each other and with external objects, while their actions can be "rewarded" (reinforced) or "punished" (inhibited) depending on the benefits they bring to the agents. For example, it is to the benefit of an agent to find sufficient food and water, and to avoid predators and obstacles in its environment. By experimenting with different rules to guide agent behavior [**Gershenson**, 2003, 2004], we expect to create a self-organizing dynamics, in which the agents come to cooperate in a coordinated system.

We start with a group of agents that are individually recognizable by "tags" or "markers" [cf. Riolo, Cohen & Axelrod, 2001; Hales & Edmonds, 2003]. The agents interact according to a game protocol with the following moves: an agent makes a request towards another agent and the other one either responds or not. Agents learn from these interactions in the following manner: if the result is positive, the agent will get more trust in the other agent's cooperativeness. Thus, the probability increases that it will make further requests to that agent in the future, or react positively to the other's requests. Vice-versa, a negative result will lead to more "distrust" and a reduced probability to make or accept requests to/from this agent.

Still, to recognise this agent, it has to take its clue from the tag, which is in general not uniquely identifiable. This means that a later interaction may be initiated with a different agent that carries a similar tag, but that is not necessarily willing to cooperate to the same extent. We may assume that if the first few interactions with agents having similar tags all generate positive (negative) results, the agent will develop a default propensity to always react positively (negatively) to agents characterised by that type of markers, while, vice-versa, the others will learn to react in the same way to the first agent.

We expect that in this way, through positive feedback, the initially undirected interactions will differentiate into a structured network of cooperative relations, in which agents with certain tags preferentially interact with agents with certain (similar or different) tags, while being reluctant to interact with others. The tags and their learned associations thus develop the function of a distributed *mediator* [**Heylighen**, 2004] that increases the probability of positive interactions by creating a differentiation between "friends" and "strangers".

In the next simulation we try to evolve a mediator that provides the group with a form of distributed cognition, i.e. an organization that allows the agents to collectively solve problems that are too complex to be tackled individually. These problems are represented as a complex of tasks. The tasks are mutually dependent in the sense that a certain task or certain tasks have to be completed before another task can be initiated. Each agent can either execute a task itself, or delegate (forward) it to another agent.

Initially all agents are equally competent or incompetent, meaning that they have the same probability of successfully accomplishing a task. However, each time it accomplishes a task, an agent becomes more "experienced" so that the probability increases that it will bring the same task to a successful end later on. We moreover assume that the agent who delegated a task will increase its trust in the competence of the agent that accomplished that task, and thus increase its probability to delegate a similar task to the same agent in the future. Otherwise, it will reduce its trust. As demonstrated by the simulation of [Gaines, 1994], this assumption is sufficient to evolve a self-reinforcing division of labour where tasks are delegated to the most "expert" agents.

However, when the tasks are mutually dependent, selecting the right specialist to carry out a task is not sufficient: First the prerequisite tasks have to be done by the right agents, in the right order. When the agents do not know a priori what the right order is, they can randomly attempt to execute or delegate a task, and, if this fails, pick out another task. Eventually they will find a task they can execute, either because it requires no preparation, or because a prerequisite task has already been done by another agent. In this way the overall problem will eventually be solved. In each problem cycle, agents will learn better when to take on which task by themselves, or when to delegate it to a specific other agent. We expect that this learned organisation will eventually stabilise into a system of efficient, coordinated actions, adapted to the task structure. While no single agent knows how to tackle the entire problem, the knowledge has been *distributed* across the system, by means of the learned associations between a tag and the competence for a particular task.

For both models our research will consist in registering and analysing the dynamics of the process of social organization as accurately as possible, by comparing the structures that emerge during the different stages of the process. In addition, different variations of the model will be tested, inspired by alternative theoretical hypotheses coming from the literature or from our own research, and by the results of preceding simulations. Specific properties that will be varied are the numbers of agents, forms of interaction (cooperation, indifference and/or conflict), strength and dynamics of trust relationships, task structure (complexity, mutual dependency), and tag distributions (fixed or variable, random or dependent on previous interactions, more or less homogeneous). This will allow us to better understand which factors contribute to an efficient organisation, and which will rather increase the risk of conflicts, fragmentation, or prejudice.

In the final stage, if we have developed a successful MAS model for the process of distributed self-organization of problem-solving, we can try to test it further with an experiment involving a group of real subjects, who are a given a complex of tasks together with "rules of the game" that are abstracted from our simulation. This will allow us to check whether the model has not overlooked any features of human interaction that essentially affect the self-organizing dynamics.

8.2. the organization co-opts external media for information sharing

To test and elaborate our second hypothesis, we need to extend our MAS with a physical environment containing virtual "objects" that can be used to permanently or temporarily store information, and thus potentially form a medium for communication between the agents. This implies that an agent should be able to change the state of an object, so that it can leave tags or markers in its environment that may later be interpreted as a signal by the same or other agents. However, if we want to understand the *self-organization* of media use, we should not assume a priori that the tags have a cognitive or communicative function. Initially, they should be seen as not more than "side effects" of the agents' actions—the way the erosion of a path is a side effect of frequent walking.

This can be achieved by having all agent actions (e.g. moving, eating, drinking, ...) leave some kind of *traces* in the shared environment. Some of these traces will be indicative of important phenomena (e.g. the proximity of food), others not. Some of the traces may remain for a long time, others will quickly be erased by changes in the environment or other agent activities. Like in the original KEBA simulation [**Gershenson**, 2002b], we assume that agents can perceive basic features of their environment (including other agents' traces), and that they learn to associate these features with other features and with their in-built goals (e.g. finding food), using classic reinforcement learning algorithms. They thus will learn to recognize which traces provide useful information about the phenomena that are important to them (e.g. food).

There seem to be two basic possibilities:

1) The trace is useful to the agent that perceives it (e.g. pointing a predator towards its prey), but detrimental to the one that made it (e.g. making the prey more visible for the predator). In that case we can expect an arms-race type of evolution, in which "predators" become better at detecting traces, while "prey" agents become better at hiding their traces. This is unlikely to lead to any kind of shared medium.
2) The trace is useful to both parties (for example because it indicates a shared danger). In this case, there will be a selective pressure for both parties to make the trace easier to perceive, by becoming more adept a leaving clear, stable and informative traces and at distinguishing and interpreting traces left by others. Thus, the trace will co-evolve with the agents' cognitive abilities, to become an efficient, shared communication medium that allows one agent to leave messages for itself and others.

To explore the ramifications of this simple model, we need to combine it with the previous simulation models in which agents learn to cooperate and coordinate. Clearly, the more efficient the pattern of cooperation that has evolved, the more useful shared media can become, and therefore the stronger the selective pressure to produce and interpret traces. Vice-versa, the better the quality of the available media, the easier it will be to evolve a sophisticated cooperative organization. Thus, we can expect that an integration of the tracing model with the self-organization model will evolve more quickly than either simulation on its own. By varying the different parameters of the model (e.g. durability of traces, sensitivity of the environment to agent activities, and sensitivity of agents to environmental features), we can try to determine the optimal combination for efficiently evolving a distributed cognitive system.

The tracing simulation unfortunately does not have an obvious analogue in human experiments, since people already start out with strong preconceptions about what constitutes a meaningful signal, and thus are unlikely to pay much attention to mere "side effects" of other people's activities (at least in the time span of a typical experiment). A more realistic set-up may offer participants the choice between direct communication (e.g. by talking) and the use of one or more indirect media (e.g. paper to jot down notes, or a shared "blackboard" on a computer system). Some media may be more helpful for certain interactions (e.g. paper to draw diagrams), and other media for others (e.g. talking to express emotions). By giving the group a complex task that requires different kinds of cognitive and communicative actions, we provide an incentive for them to self-organize, and create a division of labour—not only between individuals, but between media.

A review of the literature, theoretical analysis and the tracing simulation may give us some hints on the features of tasks and media (e.g. reliability of storage, ease of changing, ease of sharing...) that determine which kind of medium will preferentially be used for which kind of task, and how this will influence the efficiency of the distributed cognitive process. Experiments will then allow us to test these hypotheses. Moreover, we can repeat the same experiment with and without external media, to check in how far media use makes the group more effective in solving the problems posed to it. These experiments are quite innovative in psychology, where the role of media in group decision and action has rarely been studied. The approach can also be extended and embedded in the group communication experiments described in sections 8.1 and 8.4.

8.3. distributed cognitive systems function like connectionist networks

Our third hypothesis will be elaborated first by theoretical analysis examining in depth the correspondences and differences between connectionist networks and general multi-agents systems that exhibit distributed cognition, such as those proposed in the previous simulations. If possible, we would like to determine the precise conditions under which the two representations can be proven to be mathematically equivalent. Second, we will develop a recurrent connectionist model of a distributed cognitive system [cf. **Van Overwalle**, **Heylighen** & **Heath**, 2004], so that we can compare the concrete performance of the two types of models [cf. **Gershenson**, 2003; 2004].

Initially, it seems that a connectionist model may offer the following advantage. While the previous MAS models focus on the concrete *behavior* of agents, they pay little attention to the abstract information transmitted between members of a group. While this is a desirable characteristic to model the beginning stages of social self-organization in animal and human evolution, among adults collaboration is usually supported by intelligent conversation that does not focus on behavior, but on the exchange of ideas and opinions in order to coordinate collective beliefs. These collective beliefs may not have immediate implications for action, but may later on support group decisions.

To model the communication of ideas and beliefs, we make use of a standard connectionist modeling approach that has served us well in the past to model the formation and change of individual impressions, opinions and beliefs, and will extend this for a communication setting in which several individuals exchange their beliefs. We will base this approach on a standard recurrent connectionist network, which is distinguished by (a) its architecture, (b) the manner in which information is processed and (c) its learning algorithm.

(a) In a recurrent architecture, all nodes within an agent are interconnected with all of the other nodes of the same agent. Thus, all nodes send out and receive activation.

(b) Received information is represented by *external activation*, which is automatically spread among all interconnected nodes within an agent in proportion to the weights of their interconnections. The activation coming from the other nodes within an agent is called the *internal activation*.

(c) The short-term activations are stored in long-term *weight changes* of the connections; these are driven by the difference between the internal activation received from other nodes in the network and the external activation received from outside sources.

This standard recurrent model for the cognitive processes within an agent can now be extended to communication between agents [**Van Overwalle**, **Heylighen** & **Heath**, 2004], under the assumption that information is represented in broadly the same manner in different agents. Communication is represented by transferring the activation of nodes expressed by “talking” agents to “listening” agents. This is accomplished by activation spreading between agents in much the same way as activation spreading within the mind of a single agent, with the restriction that activation spreading between agents is (a) limited to nodes representing identical attributes and (b) in proportion to the connection weights linking the attributes between agents.

A crucial aspect of this between-agents dissemination of information is *trust*, or the degree to which the information on a given attribute or concept by a given agent is deemed reliable and valid. Because agents can play the role of speaker or listener, the trust connections in the model go in two directions for each agent: Sending connections for a speaking agent and receiving connections for a listening agent.

Communication is more effective if the information is believed to be trustworthy. This is implemented in the trust connection from an agent expressing its ideas to the receiving agent. When trust is maximal (+1), the information expressed by the talking agent is accepted as such by the listening agent. When trust is lower, information processing by the listener is attenuated in proportion to the trust weight. When trust is minimal (0), no information is processed by the listening agent. Thus, the listener sums all information received from talking agents in proportion to the respective trust weights, and then processes this information internally.

The criterion of novelty (section 6.4) suggests that communicators transmit only information that adds to the audience’s knowledge. On the other hand, research on group minority suggests that communicators tend to increase their interaction with an audience that does not agree with their position. This is implemented in the model by the trust weights from the listening agent to the talking agent. These weights indicate the degree of trust by the talking agent in the listening agent, and are the result of earlier communications in which the listening agent expressed judgments on an issue that were congruent with the talking agent’s knowledge. When these trust weights are high, consensual knowledge on an issue is assumed and the talking agent will refrain from expressing these ideas further. In contrast, when these weights are low, the talking agent tends to express its ideas on this issue more strongly.

Like in the standard delta learning algorithm which is used to adjust memory traces within individual agents, the degree of trust depends on the error between external beliefs expressed by a talking agent and a listening agent’s own internal beliefs. If the error is below some *trust threshold*, the trust weight between the concepts held by the two agents is increased towards 1; otherwise, the trust weight is decreased towards 0.

A distributed cognitive process is initiated when one or more agents receive one or more pieces of external information. These pieces of information may complement or confirm each other, or they may be inconsistent. The agents propagate their interpretation to “listening” agents, according to the trust connections. A listening agent will aggregate and process the information it receives from one or more talking agents. It will then pass on its own interpretation, taking into account the knowledge stored in its internal connectionist network that is the result of previous learning episodes, to others. These will again transmit their own interpretation of all the information received, in parallel or in sequence, to the other agents, and so on. At each transmission stage, the pattern of spreading activation undergoes a transformation determined by the connection pattern within and between agents, during which some information is irreversibly lost, until the connectionist network settles into an attractor. This equilibrium activation can be seen as the final, collective interpretation of the externally received information.

Using this model, we will explore a variety of configurations where different agents receive different pieces of information at different times. We will in particular investigate the effects of parameters such as the number of agents, the number of nodes per agent, the topology of the communication network (e.g. serial vs. parallel vs. circular transmission), the amount of consistent or inconsistent information that is provided as input to the system, and the distribution of this information over the agents (e.g. information can be given to one or a few agents at a time, or to the whole group) and over time (e.g. all information is provided at the beginning, or new information is added at each stage). This may allow us to already model the effect of the main factors determining the selectivity of information processing (see section 6.4), that form the focus of the next subproject (8.4).

Ultimately, we will attempt to integrate the different simulation approaches (8.1, 8.2 and 8.3) in a single model of collective acting (MAS and media) and communication (recurrent trust model). We believe that such combined model will have the most power to describe and accurately predict the different aspects of distributed cognition that we study in this project.

8.4. information in the network is propagated selectively

At all levels of communication, from the basic level of animals sharing traces to more intelligent information exchange between humans, we expect information propagation to be selective, being shaped by the needs and goals of the collective. However, before assuming that this selection is *a priori* goal-directed, we have to ask the question to what extent information propagation between humans is selective simply because of inherent features of the act of communication itself. In this respect, the interpersonal distribution of information is crucial. As noted earlier, there are two main ways in which information can be spread among people: Parallel and serial. These two communication strategies will be adopted in our experiments.

In a parallel communication design, the information is propagated from all communicators directly to the participant. In an experiment, this can be most easily controlled by using artificial communicators. For example, in a typical experiment with artificial communicators, about 80 participants read on a PC screen a number of behaviors exhibited by members of one or more target groups, presented by different communicators (e.g., “Annie saw that F., member of group A, helped an old lady across the street”). Pictures of the communicators can be used to enhance the reality of the experimental situation. The information can be consistent or inconsistent with what participants have learned earlier about the target group. Afterwards, the participants provide their own impressions about the target groups, and recall the members’ behaviors. These are the main dependent variables.

Alternatively, information exchange between real participants (3 to 4 in each group, with about 30 groups in total or 120 participants overall) can be studied, and some control over the content of the communication can be achieved by providing relevant background information before the exchange starts. During the exchange, we infuse new information in the group that needs to be discussed (e.g., in order to achieve consensus, to reach some task goal). We then measure to what extent this information is shared among the members during the open discussion, and afterward we also measure how this influenced members' opinions and recall.

A crucial issue is the distribution of information. For instance, the information can be *blocked/unshared* (the information pertaining on a given group or dimension of features is communicated by a single communicator) or *dispersed/shared* (the information is spread equally or randomly between communicators). Based on earlier simulation work [**Van Rooy, Van Overwalle** et al., 2003], we expect that different distributions (blocked vs. dispersed) will produce substantial differences in group perceptions. Other types of distribution can also be considered, such as the presentation of information either at the beginning or at the end of the communication (see also simulation work by [**Van Overwalle** & Labiouse, 2004].

Of the other factors that impact on group biases, we consider the *trustworthiness* of (the information provided by) the communicators as theoretically most crucial in this interpersonal context. This can be studied e.g. by varying whether the communicators are members of the groups or not, by varying prior expectations about the communicators, and so on. In addition, the aforementioned criteria of *utility*, *novelty*, *consistency*, *simplicity*, *expressivity* and *consensus* can be controlled to check in how far they affect the propagation.

In a serial reproduction design, the communication of information is passed from one person to another through a chain of people. For a typical experiment, there are about 30 4-participant groups. The first communicator in the chain reads the information first, and then communicates this information to the second person in the chain, and so on (e.g., "F., member of group A, helped an old lady across the street"). The information again contains consistent and inconsistent information relevant to a target group. The main dependent variables are the participant's own impression of the target groups, and memory and communication of the target member's behaviors.

Again, of main interest is the distribution of information, although this is more constrained by the serial nature of the communication. However, even in this more contrived situation, the information can be blocked (the information pertaining to a given feature is communicated first or last) or dispersed (the information is spread equally or randomly in time). We again expect that the type of distribution (blocked or dispersed) of information will lead to substantial differences in group perceptions. Also other factors pertaining to the trustworthiness of the communicators and the general selection criteria mentioned earlier will be manipulated.

From previous simulations [**Van Overwalle, Heylighen** & **Heath**, 2004; **Van Rooy, Van Overwalle** et al., 2003] and prior experiments [e.g., Lyons & Kashima, 2003], we expect that in general group impressions will become increasingly stereotypical as they are communicated by more communicators in parallel, or further down along the communication chain. However, under some circumstances, we may expect the opposite to occur. For instance, Fiedler [1991] documented that when the information is blocked as opposed to distributed, we might expect a decrease of illusory correlation biases. Although perhaps intuitively less plausible, this deviating prediction can be easily explained from a connectionist framework as developed by [**Van Rooy, Van Overwalle** et al., 2003]. This provides an interesting opportunity to study in finer detail the conditions under which group biases and stereotypes vary.

Our preliminary simulations [**Van Overwalle, Heylighen** & **Heath**, 2004] of the extended recurrent model described in the previous section also make some specific predictions with respect to the development and deployment of trust weights.

First, our simulations lead us to expect that trust between individuals is developed and applied automatically, outside of consciousness, rather than being a deliberate, controlled process. To test that the use of trust weights is automatic, we can make use of an experimental paradigm on spontaneous inferences that our research group has used before (see [**Van Overwalle**, Drenth & Marsman, 1999]). In short, in these experiments we will compare statements by trusted and distrusted sources, and see to what extent their information is spontaneously integrated in the inferences about the target. For instance, we can provide information implying some trait about the actor (e.g., the sentence "Jaana solved the mystery halfway the book" implies that Jaana is intelligent), and see to what extent this trait is also spontaneously believed by the receiving individual. We expect that this will be more the case for trusted sources than for distrusted sources, demonstrating that trust is automatically applied.

Second, we predict that speakers will spontaneously refrain from telling information that the listener already knows (the novelty criterion). We can test this by using a similar paradigm. After providing general background information on a group, we tell our participants a specific story about a member of the group and inform them that this information will have to be communicated to someone else who either does or does not possess the same background information. Immediately after the communication instruction or after telling the story, we can measure how spontaneously participants think about novel information (i.e., information inconsistent with the group's background knowledge) rather than known or consistent information. Our prediction is that when participants expect that the listener does not have sufficient background knowledge, they will activate more known and consistent story elements than novel and inconsistent elements, in preparation of the information they have to pass on. These spontaneous thoughts on story elements are measured in the same manner as above [see **Van Overwalle**, Drenth & Marsman, 1999]. Nevertheless, we expect that attenuation of talking about known information can be more easily overruled by controlled processes, such as task instructions and goals, since the act of speaking itself is largely within the control of the individual.

8.5. novel knowledge emerges

To elaborate and test our final hypothesis—that distributed cognitive systems are able to produce qualitatively new knowledge structures—we could use the integrated multiagent-connectionist simulation coming out of the first three subprojects to check in how far it produces novel concepts and relations between concepts. Alternatively, we could collaborate with our colleagues at the VUB AI-lab [e.g. Belpaeme, 2001] to extend their MAS models of the evolution of shared categories and language so as to more explicitly support distributed cognition. However, since this is the hypothesis where potentially we can expect the biggest surprises, we prefer to start with an open-ended observation of real group processes, so that it can give us a better idea of what kind of novelty can actually appear, and which factors stimulate or inhibit this form of "social construction", "collective creativity" or "distributed imagination".

To allow a quantitative analysis of our observations, we propose the following simple operationalization of knowledge creation. First, we operationalize a concept as a process of categorization, whereby different phenomena are classified as instances of this concept to a greater or lesser degree. The colour of blood, for example, will be classified with certainty (strength 1) as "red"; that of a brick - with a strength 0.7; that of an orange - with a strength 0.3; that of grass with 0. A concept can thus be represented as a vector, e.g. (1, 0.7, 0.3, 0), the components of which correspond to the categorisation strengths. Such representations in multidimensional vector spaces have proven their usefulness in the semantic analysis of concepts [**Heylighen**, 2001b; Foltz, 1996]. Then we operationalize a connection between concepts as the subjective probability or expectancy of category *A* (e.g. the phenomenon is *yellow*), given category *B* (e.g. the phenomenon is a *banana*). This determines a matrix of cross-associations between concepts [**Heylighen**, 2001b].

We can now apply these measures at both the individual and group level. For each participant, each concept is represented by a vector. The comparison of the vectors for different individuals in the group gives us an objective measure for the spread or diversity in the initial viewpoints. The average of all individual vectors defines the "collective" concept for the group [**Heylighen**, 1999]. Similarly, the average of expectancy values determines the collective association between concepts [cf. **Bollen**, 2001, 2000; **Heylighen** & **Bollen**, 2002]. After the participants have interacted, individual and collective concepts and connections can be measured again. By comparing the results before and after the group discussion, we can numerically estimate the cognitive changes that occurred in the group.

Based on our hypothesis we expect the following to hold true:

1) the spread among the participants will diminish, as exchange of information between individuals strengthens consensus;
2) the collective concept will undergo a non-linear transformation, meaning that it is no longer a linear combination of the original individual concepts:
    a) we expect that in general vector components about which there was a relative agreement will be strengthened because of conformity pressure, while components important to only one or a few individuals are suppressed, or disappear altogether.
    b) in the more specific case of novelty emergence (reaching a new attractor) we expect that vector components will be strengthened that didn't have strong values in any of the individual concepts.

These hypotheses will be tested and developed into a more detailed model by investigating the factors that control the process. At least the following factors are likely to be relevant:

diversity among the participants, *uniqueness* of their perspectives, type of *interaction*, *generality* or *ambiguity* of the concepts. A better understanding of these elements and their causal effect will allow us to choose them in such a way as to maximise the quality of the consensual concept.

In our basic set-up, a small group (about 10) of experimental participants are requested to discuss a given concept, with the objective of achieving a shared understanding. The concept is chosen such that everyone has some experience with it, but there remains sufficient vagueness or ambiguity to allow different interpretations. To minimize the risk for emotional arguments or political games, the concepts are selected to be as neutral as possible (e.g. "system", "idea", "fruit"), and the participants are told explicitly that there won't be any "winners" or "losers". The participants are informed about the concept before the experiment, so that they can prepare their thoughts without mutually influencing each other. They are asked in particular to suggest for the concept (e.g. *fruit*) a number of examples (e.g. *apple*), counterexamples (e.g. *potato*) and intermediate cases (e.g. *pumpkin*) of the category. We select the most representative ones of those, and submit the resulting list of some thirty items to all participants. We ask them to score each one on a 10-point scale, indicating the degree to which they consider it to belong to the category. This produces the initial concept vectors for all participants.

In the group discussion, each participant starts with a short description of what the concept means for him or her, and then is allowed to reply to the interpretations of others, using examples, arguments and counterarguments. After a period long enough to allow each participant to intervene several times, the discussion is stopped, and the concept vectors are measured again. The statistical comparison of initial and final vectors provides us with a quantitative analysis of the evolution of the concept. An example of novelty creation would be that after discussing it the group concludes that a *tomato* is a *fruit*, even though initially none of the participants considered it to belong to that category. A content analysis of the different interventions provides us with a more qualitative picture of the arguments and factors that have influenced the outcome. The discussion is recorded on videotape, and analysed for specific factors that appear to have influenced the outcome. The possible reasons why a particular participant has or has not changed positions are explored by focused interviews.

Complementary to this controlled experiment, we will also observe a "wild" type of discussion, using computer-mediation to record accurate data. The goal of the Principia Cybernetica Project [Heylighen, 2000; Heylighen & Joslyn, 1993] is to let a variety of experts develop a consensual theoretical framework by means of computer-supported discussion of concepts and principles. This discussion has been on-going since 1991 using electronic mail discussion lists, face-to-face meetings, and the web [Heylighen, Joslyn & Turchin, 1993-2004]. There is plenty of textual material available recording past discussions, which can be analysed to look for the novelty-creating processes that we hypothesize. Moreover, by providing the participants with a more structured computer-mediation, such as the CLAIMAKER argumentation environment developed by a group associated with Principia Cybernetica [Shum et al., 2003], we can accurately register the different "moves" in future, open-ended discussions within the group. By moreover asking participants to score the connections between the concepts that are likely to be discussed before and after the extended discussion (which can last months), we get a quantitative measure of the changes.

In addition, we already have access to many hours of video-tape recordings of other, non-controlled group discussions (e.g. during the Global Brain Workshop) where participants were similarly attempting to develop novel, consensual insights. These too will be checked for the hypothesized processes and monitored for unexpected phenomena.

9. Deliverables

In addition to the novel insights and conceptual framework, we expect this project to deliver the following more concrete "products".

9.1. Publications

At the end of the 5 year duration of this project, we expect to have published dozens of papers with the results of our research, both theoretical and empirical, in a variety of international, peer-refereed journals, as well as in proceedings of conferences, and as chapters in books. Moreover, we plan to write at least two monographs:

1) a textbook for researchers and advanced students formally elaborating a conceptual framework for the modelling of distributed cognitive systems and their evolution. Following a similar structure as this proposal, it will start with the most simple element (objects, interactions, agents), and show how these can self-organize step by step to produce gradually more complex systems (groups, division of labor, distributed problem-solving, learning, coordination, etc.). The general principles will be illustrated with concrete examples, such as insect societies, organizations, group processes, socio-cultural evolution, or the coordination of software agents.
2) a practical handbook with exercises showing how to model distributed cognitive systems using our generic connectionist simulation environment (see below).

In addition, this project should produce at least three PhD dissertations, investigating different computational and empirical aspects of our general project.

9.2. Simulation environments

Many of the connectionist simulations will be conducted with the aid of a software program, called FIT, developed by **Van Overwalle**. Since several years, this program is widely available on the Internet: www.vub.ac.be/PESP/VanOverwalle.html#fit
As the program is extended with the results of our research, more advanced versions will also be made freely available to the research community.

Similarly, the KEBA multi-agent simulation environment [Gershenson, 2002] as it is extended with social self-organization and media sharing, will also be made available on the Internet, as will be an eventual integrated environment, combining the strengths of connectionist and multi-agent approaches. In this way, other researchers and students will be able not only to replicate our results, but to devise their own models of distributed cognitive systems and explore their properties in a flexible and user-friendly manner.

9.3. Empirical data

Like the software we develop, we also plan to make the data gathered from the experiments and observations available via web, so that other researchers can use them to re-analyse and to test their own hypotheses.

9.4. Workshops, conferences and lectures

Like in the past, we will continue to regularly organize international meetings on the subject of distributed cognition and its specific aspects, so that our work can be discussed with other researchers in the domain, and receive input from their results. The talks presented at the more important meetings will be published in the form of proceedings. We will also present our ideas in seminars and lectures for local colleagues and PhD students, and include the most important insights in the undergraduate courses we teach.

10. Project planning

The research project is scheduled to run for 5 years, from Jan. 1, 2005 to Dec. 31, 2009.

Year 1: 2005

In the first year, we will start with the two subprojects (8.4 and 8.5) that center around laboratory experiments, since these are most likely to be time-consuming, while running the greatest risk of failure, so that initial experiments may need to be redone or redesigned. For each of these two empirical projects we will need to employ a new research assistant (by means of a 4-year PhD scholarship) with a social science background, to set up and run the experiment and process the data. In the meantime, the present members of the team will focus on the literature review, theoretical analysis and preliminary connectionist simulations, so as to put the conceptual framework on a firm foundation, while providing guidelines for the design of the experiments.

Year 2: 2006

In the second year, while the experiments and the theoretical and connectionist modelling are running and , we will set up the more complex MAS simulations that form the core of subprojects 8.1 and 8.2, building on the preliminary theoretical and empirical results. This will require the employment of another research assistant, with extensive computing experience, to program the simulations, run the different variations, collect and process the data. In this year, we also plan to organize a first project workshop with all team members and invited outside experts, to discuss the first results.

Year 3: 2007

After two years of empirical data collection and connectionist simulation, and one year of MAS simulation, we will have sufficient material to start developing an integrated theoretical and simulation platform that combines MAS and connectionist principles (see subproject 8.3). This will require a fourth, more experienced researcher, at the PostDoc level. This researcher will keep close contact with the on-going experiments and agent simulations, to use their insights to build the integrated platform, and to suggest additional variations for testing. We will further run another project workshop to keep all people involved up-to-date about the advances and as yet unresolved issues.

Year 4: 2008

After three years of experiments, the two initial research assistants should have collected sufficient data to analyse and draw general conclusions so that they can defend their PhD dissertations on the subject by the end of the year. The simulations will continue to run different variations, while being extended with new insights and hypotheses coming from the experiments and theoretical investigations. A third international workshop is organized.

Year 5: 2009

After three years of agent simulations, the third research assistant too will have collected sufficient data to analyze and interpret in the form of a PhD dissertation. The PostDoc researcher will complete the development and data processing of the integrated platform. We conclude the project with a large, international conference on the broad subject of distributed cognition, with both invited and submitted papers from specialists around the world, during which the members of our team present all the major results of the project to the academic community.

Requested Funding

The following budget (all costs in Euro) provides an estimate of the funding we will need over the 5 years to run the project, split up into the different cost categories.

YEAR	2005	2006	2007	2008	2009	TOTAL
Purchase of books and journals	4000	4000	4000	4000	4000	20000
Travel and accomodation, to let team members participate in scientific conferences and visit research centers abroad	8000	8000	8000	8000	8000	40000
Organization of workshops and travel+ accommodation for visiting experts	10000	10000	10000	10000	10000	50000
Payment of participants in experiments (800 people x 7.5 euro/person)	6000	6000	6000	6000	6000	30000
Scientific Software licences (Statistics, experiment generator, development platforms...)	4000	4000	4000	4000	4000	20000
Various Scientific Material	1000	1000	1000	1000	1000	5000
computer equipment for PhD student 1	2500					2500
computer equipment for PhD student 2	2500					2500
computer equipment for PhD student 3		2500				2500
computer equipment for PostDoc 4			2500			2500
PhD scholarship 1	28979	29559	31737	32272		122547
PhD scholarship 2	28979	29559	31737	32272		122547
PhD scholarship 3		29559	31737	32272	32917	126485
PostDoc contract 4			61583	64296	70852	196731
TOTALS	95958	124177	192294	194112	136769	**743310**

Relevant publications of the research team

The following is a selection of the most important publications of the research team that are relevant for the present proposal.

Bollen. J. (2000) Group User Models for Personalized Hyperlink Recommendations. In LNCS 1892 *International Conference on Adaptive Hypermedia and Adaptive Webbased Systems* (AH2000), pages 39-50, Trento, August . Springer Verlag.

Bollen J. (2001) *A Cognitive Model of Adaptive Web Design and Navigation - A Shared Knowledge Perspective*, Free University of Brussels, Faculty of Psychology, PhD Dissertation.

Bollen J. & **Heylighen** F. (1996) "Algorithms for the Self-organisation of Distributed, Multi-user Networks", in: *Cybernetics and Systems '96* R. Trappl (ed.), (Austrian Society for Cybernetics), p. 911-916.

Bollen J., **Heylighen** F. (1998): "A system to restructure hypertext networks into valid user models", *New Review of HyperMedia and Multimedia* 4, p. 189-213.

Bollen, J., **Heylighen** F., **Van Rooy** D. (1998): "Improving Memetic Evolution in Hypertext and the WWW", in: *Proc. 16th Int. Congress on Cybernetics* (Association Internat. de Cybernétique, Namur), p. 449-454.

Chielens, K. (2003) *The Viral Aspects of Language: A Quantitative Research of Memetic Selection Criteria*. Unpublished Masters Thesis VUB.

Gershenson, C. (2001). *Artificial Societies of Intelligent Agents*. Unpublished BEng Thesis. Fundacion Arturo Rosenblueth, Mexico.

Gershenson, C. (2002a). Philosophical Ideas on the Simulation of Social Behaviour. *Journal of Artificial Societies and Social Simulation* vol. 5, no. 3.

Gershenson, C. (2002b). Behaviour-based Knowledge Systems: An Epigenetic Path from Behaviour to Knowledge. *Proceedings of the 2nd Workshop on Epigenetic Robotics*. Edinburgh.

Gershenson, C. (2003). Comparing Different Cognitive Paradigms with a Virtual Laboratory. IJCAI-03: *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pp. 1635-6. Morgan Kaufmann.

Gershenson, C. (2004). Cognitive Paradigms: Which One is the Best? *Cognitive Systems Research*, 5(2):135-156, June 2004.

Gershenson, C., F. **Heylighen** (2003). "When Can we Call a System Self-organizing?", In Banzhaf, W, T. Christaller, P. Dittrich, J. T. Kim and J. Ziegler (eds.), Advances in Artificial Life, 7th European Conference, ECAL 2003, (Springer, LNAI 2801.), p. 606-614.

Gershenson, C., F. **Heylighen** (2004a). How can we think the complex? in: Richardson, Kurt (ed.) *Managing the Complex Vol. 1: Philosophy, Theory and Application.* [in press]

Gershenson C., **Heylighen** F. (2004b): Protocol Requirements for Self-organizing Artifacts: Towards an Ambient Intelligence, in: *Proc. Int. Conf. on Complex Systems* (New England Institute of Complex Systems)

Gershenson, C., M. A. Porter, A. Probst, M. Marko and A. Das. (2002) A Study on the Relevance of Information in Discriminative and Non-Discriminative Media. *InterJournal of Complex Systems* , 533.

Gershenson, C., P. P. González and J. Negrete. (2000a) Action Selection Properties in a Software Simulated Agent, in Cairó et. al. (Eds.) *MICAI 2000: Advances in Artificial Intelligence*. Lecture Notes in Artificial Intelligence 1793, pp. 634-648. Springer-Verlag.

Gershenson, C., P. P. González and J. Negrete. (2000b) Thinking Adaptive: Towards a Behaviours Virtual Laboratory. In Meyer et. al. (eds.) *Simulation of Adaptive Behavior 2000 Proceedings Supplement*. Paris, France. ISAB press.

Heath M. (in preparation): *The Possibility of Radical Novelty in Evolving Cognitive Systems* (PhD thesis, Dept. of Psychology, Vrije Universiteit Brussel)

Heylighen F. (1989): "Causality as Distinction Conservation: a theory of predictability, reversibility and time order", *Cybernetics and Systems* 20, p. 361-384.

Heylighen F. (1989): “Self-Organization, Emergence and the Architecture of Complexity”, in: *Proc. 1st European Conference on System Science*, (AFCET, Paris), p. 23-32.

Heylighen F. (1990a): “Autonomy and Cognition as the Maintenance and Processing of Distinctions”, in: *Self-Steering and Cognition in Complex Systems*, **Heylighen** F., Rosseel E., Demeyere F. (ed.), (Gordon and Breach, New York), p. 89-106.

Heylighen F. (1990b): *Representation and Change. A Metarepresentational Framework for the Foundations of Physical and Cognitive Science*, (Communication and Cognition, Gent), 200 p.

Heylighen F. (1990c): “A New Transdisciplinary Paradigm for the Study of Complex Systems?”, in: *Self-Steering and Cognition in Complex Systems*, **Heylighen** F., Rosseel E. & Demeyere F. (ed.), (Gordon and Breach, New York), p. 1-16.

Heylighen F. (1991): “Cognitive Levels of Evolution: pre-rational to meta-rational”, in: *The Cybernetics of Complex Systems - Self-organization, Evolution and Social Change*, F. Geyer (ed.), (Intersystems, Salinas, California), p. 75-91.

Heylighen F. (1991): "Structuring Knowledge in a Network of Concepts", in : *Workbook of the 1st Principia Cybernetica Workshop*, Heylighen F. (ed.) (Principia Cybernetica, Brussels-New York), p. 52-58.

Heylighen F. (1991): “Design of a Hypermedia Interface Translating between Associative and Formal Representations”, *International Journal of Man-Machine Studies* 35, p. 491-515.

Heylighen F. (1991): “Modelling Emergence”, *World Futures: the Journal of General Evolutio*n 31 (Special Issue on Emergence, edited by G. Kampis), p. 89-104.

Heylighen F. (1992): "Non-Rational Cognitive Processes as Changes of Distinctions", in: *New Perspectives on Cybernetics. Self-Organization, Autonomy and Connectionism*, G. Van de Vijver (ed.), (Synthese Library v. 220, Kluwer Academic, Dordrecht), p. 77-94.

Heylighen F. (1992) : “‘Selfish’ Memes and the Evolution of Cooperation”, *Journal of Ideas* , Vol. 2, #4, pp 77-84.

Heylighen F. (1992) : “Evolution, Selfishness and Cooperation”, *Journal of Ideas*, Vol 2, # 4, pp 70-76.

Heylighen F. (1992): “A Cognitive-Systemic Reconstruction of Maslow’s Theory of Self-Actualization”, *Behavioral Science* 37, p. 39-58.

Heylighen F. (1992): “Principles of Systems and Cybernetics: an evolutionary perspective”, in: *Cybernetics and Systems ‘92*, R. Trappl (ed.), (World Science, Singapore), p. 3-10.

Heylighen F. (1993): “Selection Criteria for the Evolution of Knowledge”, in: *Proc. 13th Int. Congress on Cybernetics* (Association Internat. de Cybernétique, Namur), p. 524-528.

Heylighen F. (1994) “Fitness as Default: the evolutionary basis for cognitive complexity reduction”, in: Cybernetics and Systems ‘94, R. Trappl (ed.), (World Science, Singapore), p.1595-1602.

Heylighen F. (1995): “(Meta)systems as constraints on variation”, *World Futures: the Journal of General Evolution* .45, p. 59-85.

Heylighen F. (1997): “Objective, subjective and intersubjective selectors of knowledge”, *Evolution and Cognition* 3:1, p. 63-67.

Heylighen F. (1997): “The Economy as a Distributed, Learning Control System”, *Communication and Cognition- AI* 13, nos. 2-3, p. 207-224.

Heylighen F. (1998): “What makes a meme successful? Selection criteria for cultural evolution”, in: *Proc. 16th Int. Congress on Cybernetics* (Association Internat. de Cybernétique, Namur), p. 423-418.

Heylighen F. (1999): “Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map”, *Computational and Mathematical Theory of Organizations* 5(3), p. 253-280.

Heylighen F. (1999): “The Growth of Structural and Functional Complexity during Evolution”, in: F. **Heylighen**, J. **Bollen** and A. Riegler (eds.) The Evolution of Complexity (Kluwer Academic, Dordrecht), p. 17-44.

Heylighen F. (2000): “Foundations and Methodology for an Evolutionary World View: a review of the Principia Cybernetica Project”, *Foundations of Science*, 5, p. 457-490.

Heylighen F. (2001a): "Bootstrapping knowledge representations: from entailment meshes via semantic nets to learning webs", *Kybernetes* 30 (5/6), p. 691-722.

Heylighen F. (2001b): "Mining Associative Meanings from the Web: from word disambiguation to the global brain", in: *Proceedings of the International Colloquium: Trends in Special Language and Language Technology*, R. Temmerman and M. Lutjeharms (eds.) (Standaard Editions, Antwerpen), p. 15-44.

Heylighen F. (2002): "The Science of Self-organization and Adaptivity", in: Knowledge Management, Organizational Intelligence and Learning and Complexity, in: *The Encyclopedia of Life Support Systems*, (Eolss Publishers, Oxford).

Heylighen F. (2004): "Mediator Evolution: a general scenario for the origin of dynamical hierarchies", Artificial Life [submitted]

Heylighen F. (2004): "The Global Superorganism: an evolutionary-cybernetic model of the emerging network society", Journal of Collective Intelligence [submitted]

Heylighen F., **Bollen** J. (1996) "The World-Wide Web as a Super-Brain: from metaphor to model", in: C*ybernetics and Systems '96* R. Trappl (ed.), (Austrian Society for Cybernetics).p. 917-922.

Heylighen F., **Bollen** J. (2002): "Hebbian Algorithms for a Digital Library Recommendation System", in *Proceedings 2002 International Conference on Parallel Processing Workshops* (IEEE Computer Society Press)

Heylighen F., **Bollen** J., Riegler A. (ed.) (1999): *The Evolution of Complexity* (Kluwer Academic, Dordrecht).

Heylighen F., Campbell D.T. (1995): "Selection of Organization at the Social Level: obstacles and facilitators of metasystem transitions", *World Futures: the Journal of General Evolution* 45, p. 181-212.

Heylighen F., Dewaele J-M. (2002): "Variation in the contextuality of language: an empirical measure", *Foundations of Science* 6, p. 293-340

Heylighen F., **Gershenson** C. (2003): "The Meaning of Self-organization in Computing", IEEE Intelligent Systems 18:4, p. 72-75.

Heylighen F., **Heath** M. (eds.) (2004): From Intelligent Networks to the Global Brain, special issue of *Technological Forecasting and Social Change* (in press)

Heylighen F., **Heath** M., F. **Van Overwalle** (2004): The Emergence of Distributed Cognition: a conceptual framework, submitted for Proceedings of Collective Intentionality IV, Siena (Italy), to be published as a special issue of *Cognitive Systems Research*

Heylighen F., Joslyn C. (1993): "Electronic Networking for Philosophical Development in the Principia Cybernetica Project", *Informatica* 17, No. 3, p. 285-293.

Heylighen F., Joslyn C. (1995): "Systems Theory", in: *The Cambridge Dictionary of Philosophy*, R. Audi (ed.) (Cambridge University Press, Cambridge), p.784-785.

Heylighen F., Joslyn C. (2001): "Cybernetics and Second Order Cybernetics", in: R.A. Meyers (ed.), *Encyclopedia of Physical Science and Technology* (3rd ed.), Vol. 4 , (Academic Press, New York), p. 155-170.

Heylighen F., Joslyn C. & Turchin V. (eds.) (1993-2004): "Principia Cybernetica Web" (http://pespmc1.vub.ac.be/),

Heylighen F., Rosseel E., Demeyere F. (eds.) (1990): *Self-Steering and Cognition in Complex Systems. Toward a New Cybernetics*, (Gordon and Breach Science Publishers, New York), 440 p.

Jordens, K., **Van Overwalle, F.** (2001). Empirische toetsing van een feedforward connectionistisch model van cognitieve dissonantie: de rol van affect in het geïnduceerd-inwilligingsparadigma [Empirical testing of a feedforward connectionist model of cognitive dissonance: the role of affect in the induced compliance paradigm.] In D. Stapel, C. Martijn, E. van Dijk,, A. Dijksterhuis (Eds.) *Fundamentele sociale psychologie* (Vol. 15, pp. 91—102). Delft, The Netherlands: Eburon.

Jordens, K., **Van Overwalle, F**. (2004). Connectionist Modeling of Attitudes and Cognitive Dissonance. In G. Haddock, G. Maio (Eds.) *Contemporary perspectives on the psychology of attitudes*. London: Psychology Press.

Jordens, K.,, **Van Overwalle, F**. (submitted) *Cognitive dissonance and affect: An empirical test of a connectionist account.*

Martens B. (1998): “The relationship between knowledge and economic growth: an economic-cognitive approach” *Proc. 23rd Flemish Economic Congress*, Leuven.

Martens B. (1999): “The introduction of complexity concepts in economics: towards a new paradigm”, in: F. **Heylighen**, J. **Bollen** and A. Riegler (eds.) *The Evolution of Complexity* (Kluwer Academic, Dordrecht),

Martens B. (1999): “Towards a Generalised Coase Theorem: a theory of the emergence of social and institutional structures under imperfect information”, in Barnett et al, (eds.): *Commerce, Complexity and Evolution*, Cambridge University Press.

Martens B. (2004): *The cognitive mechanics of economic development and social change*, (PhD Thesis, Vrije Universiteit Brussel).

Martens B., P. Murrell, P. Seabright and U. Mummert (2002), *The institutional economics of foreign aid,* Cambridge University Press.

Rocha L. M., J. **Bollen** (2000). Biologically motivated distributed designs for adaptive knowledge management. In I.Cohen and L. Segel, editors, *Design Principles for the Immune System and other Distributed Autonomous Systems*, Oxford University Press. pp. 305-334.

Timmermans, B., & Cleeremans, A. (2000). Rules versus statistics in biconditional grammar learning. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, 947-952. NJ: Erlbaum

Timmermans, B., **Van Overwalle, F.** (submitted) Spontaneous circumstantial attributions.

Van Overwalle, F. (1989). Structure of freshmen’s causal attributions for exam performance. *Journal of Educational Psychology, 81*, 400-407. (**SSCI=1.844**)

Van Overwalle, F. (1997a) Dispositional attributions require the joint methods of difference and agreement. *Personality and Social Psychology Bulletin, 23*, 974-980. (**SSCI=1.748**)

Van Overwalle, F. (1997b) A test of the Joint model of causal attribution. *European Journal of Social Psychology, 27*, 221-236. (**SSCI=0.826**)

Van Overwalle, F. (1998) Causal explanation as constraint satisfaction : A critique and a feedforward connectionist alternative. *Journal of Personality and Social Psychology, 74*, 312-328. (**SSCI=2.873**)

Van Overwalle, F. (2003) Acquisition of dispositional attributions: Effects of sample size and covariation. *European Journal of Social Psychology*, 33, 515—533. (SSCI=1.449)

Van Overwalle, F. (2004). Multiple Person Inferences: A View of a Connectionist Integration. In H. Bowman, C. Labiouse (Eds.), *Proceedings of the Eighth Neural Computation and Psychology Workshop* (Progress in Neural Processing). London, UK: World Scientific

Van Overwalle, F. (under revision) *Discounting and augmentation of dispositional and causal attributions.*

Van Overwalle, F., Drenth, T., Marsman, G. (1999). Spontaneous trait inferences : Are they linked to the actor or to the action ? *Personality and Social Psychology Bulletin, 25*, 450-462. (**SSCI=1.748**)

Van Overwalle, F., **Heylighen**, F. (1991). Invariantie-kenmerken bij antecedente condities en attributionele dimensies : Waarnemen van oorzaken, verwachtingen en emoties. [Invariance features in antecedent conditions and attributional dimensions: Perceiving causes, expectations and emotions] In J. van der Pligt, W. van der Kloot, A. van Knippenberg and M. Poppe (Eds.) *Fundamentele sociale psychologie: Deel 5* (pp. 44-60). Tilburg : Tilburg University Press.

Van Overwalle, F., **Heylighen**, F. (1995) Relating covariation information to causal dimensions through principles of contrast and invariance. *European Journal of Social Psychology*, 25, 435-455. (**SSCI=0.826**)

Van Overwalle, F., **Heylighen**, F., Casaer, S., Daniëls, M. (1992). Preattributional and attributional determinants of emotions and expectations. *European Journal of Social Psychology, 22*, 313-329. (**SSCI=0.826**)

Van Overwalle, F., **Heylighen,** F., **Heath** M. (2004): From Communication between Individuals to Collective Beliefs, submitted for Proceedings of Collective Intentionality IV, Siena (Italy), to be published as a special issue of *Cognitive Systems Research*

Van Overwalle, F., Jordens, K. (2002). An adaptive connectionist model of cognitive dissonance. *Personality and Social Psychology Review*, 3, 204—231. (**SSCI=3.222**)

Van Overwalle, F., Labiouse, C. (2004) A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review*, 8, 28—61. (**SSCI=3.222**)

Van Overwalle, F., Mervielde, I., De Schuyter, (1995). Structural modeling of the relationships between attributions, emotions and behavior of college freshmen. *Cognition and Emotio*n, 9, 59-85. (**SSCI= 1.863**)

Van Overwalle, F., Siebler, F. (submitted). A Connectionist Model of Attitude Formation and Change.

Van Overwalle, F., Timmermans, B. (2001). Learning about an Absent Cause: Discounting and Augmentation of Positively and Independently Related Causes. French, R.M.,, Sougné, J.P. (Eds.) *Connectionist Models of Learning, Development and Evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop, Liege, Belgium, 16-18 September 2000*. Springer Verlag.

Van Overwalle, F., **Timmermans**, B. (under revision) Discounting and augmentation in attribution: The role of the relationship between causes.

Van Overwalle, F., **Van Rooy**, D. (2001a). How one cause discounts or augments another: A connectionist account of causal competition. *Personality and Social Psychology Bulletin*, 27, 1613—1626. (**SSCI=1.748**)

Van Overwalle, F., **Van Rooy**, D. (2001b). When more observations are better than less : A connectionist account of the acquisition of causal strength. *European Journal of Social Psychology*, 31, 155-175. (**SSCI= 1.449**)

Van Overwalle, F.,, **Van Rooy**, D. (1998) A Connectionist Approach to Causal Attribution. In S. J. Read, L. C. Miller (Eds.) *Connectionist and PDP models of Social Reasoning and Social Behavior* (pp. 143—171). Lawrence Erlbaum.

Van Rooy, D. (2000): *A connectionist model of illusory correlation*.(PhD Thesis, Vrije Universiteit Brussel).

Van Rooy, D., **Van Overwalle, F**. (submitted) Illusory correlation, sample size and memory: A connectionist approach.

Van Rooy, D., **Van Overwalle, F.**, **Vanhoomissen**, T., Labiouse, C., French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review*, 110, 536-563. (**SSCI=6.750**)

Bibliography (publications by others)

Ashby, W. R. (1962). Principles of the Self-organizing System. In von Foerster, H. and G. W. Zopf, Jr. (Eds.), *Principles of Self-organization*. Pergamon Press, pp. 255-278.

Aunger R. (ed.) (2001): *Darwinizing Culture: The Status of Memetics As a Science* (Oxford University Press)

Axelrod, R. M., *The Evolution of Cooperation*, Basic Books New York (1984).

Belpaeme T. (2001) Reaching coherent color categories through communication. In Kröse, B. et al. (eds.), *Proc. 13th Belgium-Netherlands Conference on AI,* Amsterdam, p. 41-48.

Berger P. L., T. Luckmann: (1967) *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Anchor.

Berners-Lee ,T., J. Hendler & O. Lassila (2001): The Semantic Web, *Scientific American*, 282(5),

Bonabeau E., Dorigo M. and Theraulaz G. (1999) *Swarm intelligence: From natural to artificial systems*. Oxford University Press.

Brauer, M., Judd, C. M., & Jacquelin (2001). The communication of social stereotypes: The effects of group discussion and information distribution on stereotypic appraisals. *Journal of Personality and Social Psychology, 81*, 463—475.

Cantwell Smith, B. (1996): *On the origin of objects* (MIT Press)

Clark A. and Chalmers D. (1998):, "The Extended Mind," *Analysis* 58, p. 7-19.

Clark, A. (1997). *Being There: putting brain, body, and world together again*, Cambridge, Mass., MIT Press.

Crowston, K. (2003). A taxonomy of organizational dependencies and coordination mechanisms. In Malone, T. W., Crowston, K. and Herman, G. (Eds.) *Tools for Organizing Business Knowledge: The MIT Process Handbook*. Cambridge, MA: MIT Press.

Crutchfield J. (1998). Dynamical embodiments of computation in cognitive processes, *Behavioral and Brain Sciences*, 21, p. 635.

Crutchfield J., Shalizi C., Tumer K & Wolpert D. (eds.) (2002): *Collective Cognition Workshop Proceedings: Mathematical Foundations of Distributed Intelligence* (http://www.santafe.edu/~dynlearn/colcog/, to be published in the Santa Fe Institute Studies in the Sciences of Complexity, Oxford University Press;)

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An Information loss account of distinctiveness -based illusory correlations. Journal of Personality and Social Psychology, 60, 24-36.

Foltz P.W. (1996) Latent Semantic Analysis for text-based research, *Behavior Research Methods, Instruments, & Computers*, 28, 197-202.

Gaines B.R. (1994), The Collective Stance in Modeling Expertise in Individuals and Organizations, *Int. J. Expert Systems* 71, 22-51.

Hales, D., and B. Edmonds, "Evolving social rationality for MAS using "tags"", *Proc. 2nd Int. Conference on Autonomous Agents and Multiagent Systems* (J. S. R. et Al. ed.), ACM Press (2003), 497–503.

Hutchins E (1995): *Cognition in the Wild* (MIT Press).

Hutchins, E. & B. Hazelhurst (1995). How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert and R. Conte (Eds.), *Artificial Societies*. UCL Press

ISTAG (2003): *Ambient Intelligence: from vision to reality* (report to the European Commission, available at http://www.cordis.lu/ist/istag.htm)

Janis, I. L. (1972) *Victims of groupthink*. (Boston: Houghton Mifflin).

Kirsch, D. & Maglio, P. (1994) On distinguishing epistemic from pragmatic action. *Cognitive Science* 18: 513-549

Kirsh, D. (1996) Adapting the Environment Instead of Oneself. *Adaptive Behavior*, Vol 4, No. 3/4, 415-452.

Klein, O. Jacobs, A., Gemoets, S. Licata, L. & Lambert, S. (2003). Hidden profiles and the consensualization of social stereotypes: how information distribution affects stereotype content and sharedness. *European Journal of Social Psychology, 33*, 755—777.

Lévy P. (1997): *Collective Intelligence*, Plenum.

Lyons, A. & Kashima, Y. (2003) How Are Stereotypes Maintained Through Communication? The Influence of Stereotype Sharedness. *Journal of Personality and Social Psychology, 85*, 989-1005.

McLeod, P., Plunkett, K. & Rolls, E. T. (1998). *Introduction to connectionist modeling of cognitive processes*. Oxford, UK: Oxford University Press.

Riolo, R., M. D. Cohen, and R. M. Axelrod (2001), "Evolution of cooperation without reciprocity", *Nature* 414, 441–443.

Rumelhart D.E. & J.L. McClelland (editors) (1986): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. MIT Press.

Searle, J. (1995): *The Construction of Social Reality*, Free Press

Senge P. (1990) T*he Fifth Discipline: the Art and Practice of Learning Organizations*, Doubleday.

Shum S.B., Vi. Uren, G. Li, Jo. Domingue, E. Motta (2003): Visualizing Internetworked Argumentation, In: *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Paul A. Kirschner, Simon J. Buckingham Shum and Chad S. Carr (Eds), . Springer-Verlag: London

Staab S. & Studer R., (eds.) (2003), *Handbook on Ontologies in Information Systems*, Springer Verlag.

Steels L. (1998): Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation, in Hurford et al. (eds): *Approaches to the evolution of language* (Cambridge University Press), p. 384-404.

Steels L. & Brooks R. (eds.) (1995): *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents* (Erlbaum).

Susi, T. & Ziemke, T. (2001). Social Cognition, Artefacts, and Stigmergy: A Comparative Analysis of Theoretical Frameworks for the Understanding of Artefact-mediated Collaborative Activity. *Cognitive Systems Researc*h, 2(4), 273-290.

von Foerster, H., (1960). On self-organising systems and their environments, in: *Self-Organising Systems*, M.C. Yovits and S. Cameron (eds.), Pergamon Press, London, pp. 30-50.

Wright, R.(2000): *Non-Zero. The Logic of Human Destiny* (Pantheon Books)