2006

Agents, idols, and icons

Pascal Perez
University of Wollongong, pascal@uow.edu.au

Publication Details

Perez, P. (2006). Agents, idols, and icons. In P. Perez & D. Batten (Eds.), Complex Science for a Complex World: Exploring Human Ecosystems with Agents (pp. 27-56). Canberra, Australia: ANU E Press.

Agents, idols, and icons

Abstract

Since the early 1960s, Artificial Intelligence has cherished the ambition to design an artificial cognitive machine able to reproduce intimate aspects of human behaviour. Distributed Artificial Intelligence and its most recent avatars—Multi- Agent Systems—have developed the concept towards social interactions and societal dynamics, attracting the attention of sociologists and ethnographers who found new ways to elaborate or validate their theories. But populations of cognitive agents aren't the real thing, despite the efforts of their designers. Furthermore, one must cautiously examine the rationale behind these often incredibly complex arrangements of algorithms, in order to assess the usefulness of such exercises. As a matter of fact, Artificial Intelligence relies on a very positivist, and sometimes reductionist, view of human behaviour. For centuries, from Bacon to Pierce, philosophy of mind has provided meaningful insights that challenge some of these views. More recently, post-normal approaches have even taken a more dramatic stand—some sort of paradigm shift—where direct knowledge elicitation and processing override the traditional hardwiring of formal logic-based algorithm within computer agents. Keywords: Agent-Based Modelling, Artificial Intelligence, Icon, Idol, Philosophy of Mind, Cognition.

Keywords

idols, icons, agents

Disciplines

Engineering | Physical Sciences and Mathematics

Publication Details

Perez, P. (2006). Agents, idols, and icons. In P. Perez & D. Batten (Eds.), Complex Science for a Complex World: Exploring Human Ecosystems with Agents (pp. 27-56). Canberra, Australia: ANU E Press.

3. Agents, Icons and Idols

Pascal Perez

Abstract

Since the early 1960s, Artificial Intelligence has cherished the ambition to design an artificial cognitive machine able to reproduce intimate aspects of human behaviour. Distributed Artificial Intelligence and its most recent avatars—Multi-Agent Systems—have developed the concept towards social interactions and societal dynamics, attracting the attention of sociologists and ethnographers who found new ways to elaborate or validate their theories. But populations of cognitive agents aren't the real thing, despite the efforts of their designers. Furthermore, one must cautiously examine the rationale behind these often incredibly complex arrangements of algorithms, in order to assess the usefulness of such exercises. As a matter of fact, Artificial Intelligence relies on a very positivist, and sometimes reductionist, view of human behaviour. For centuries, from Bacon to Pierce, philosophy of mind has provided meaningful insights that challenge some of these views. More recently, post-normal approaches have even taken a more dramatic stand—some sort of paradigm shift—where direct knowledge elicitation and processing override the traditional hardwiring of formal logic-based algorithm within computer agents. *Keywords*: Agent-Based Modelling, Artificial Intelligence, Icon, Idol, Philosophy of Mind, Cognition.

Introduction

Scientists developing Multi-Agent Systems, as part of Distributed Artificial Intelligence (DAI), tend to focus on individual components interacting within a given system (Gilbert and Troitzsch 1999). This is a purely bottom-up approach where representations of the individual components, the agents, display a large autonomy of action. Hence, system-level behaviours and patterns emerge from a multitude of local interactions. Intentionality is deliberately placed at the level of the agents to the detriment of the system itself, greatly limiting its ability to control its own evolution. In the case of human ecosystems, agents can represent individual actors or relevant social groups and communities (Bousquet and LePage 2004). The following definition of a Multi-Agent System (MAS) is generally admitted. A MAS is a conceptual model of an observed system that includes:

- an environment (E), often possessing explicit metrics;
- a set of passive objects (O), eventually created, destroyed or modified by the agents;

- a set of active agents (A). Agents are autonomous and active objects of the system;
- a set of relationships (R), linking objects and/or agents together; and
- a set of operators (Op), allowing agents to perceive, create, use, or modify objects.

An agent is a physical or virtual entity that demonstrates the following abilities: autonomy, communication, limited perception, bounded rationality, and decision-making process based on satisfying goals and incoming information (Ferber 1999). A Multi-Agent Based Simulation (MABS) is the result of the implementation of an operational model (computer-based), designed from a MAS-based conceptual representation of an observed system. The strength of MAS approaches consists in their ability to represent socially and spatially distributed problems. Meaningful examples of application are to be found in ecology (Janssen 2002), sociology (Conte and Castelfranchi 1995), or economics (Tesfatsion 2001).

Cederman (2005) asserts that generative process theorists in social science, shifting from traditional nomothetic to generative explanations of social forms and from variable-based to configurative ontologies, may find in Multi-Agent Systems relevant tools to explore the emergence of social forms in the Simmelian tradition, thanks to common foundations in both epistemology and ontology.

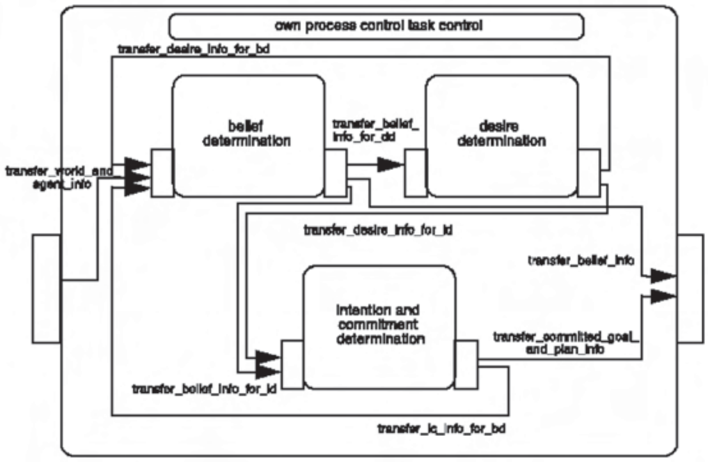
In the following sections of this chapter, we try to evaluate Cederman's assertion against evidence. First, we describe general features of cognitive agents as stated and used in Artificial Intelligence (AI) research. Then, we argue that our understanding of mental processes, from Bacon's idols to Tversky's prospect theory, is inherently limited. In the third section, we question the supposed objective autonomy of agents, drawing from Peirce's icons and Varela's enactive cognitive theory. Finally, we propose a way forward that encapsulates the designer and the modelling process into the observed system itself.

Cognitive agents

Kenetics (Ferber 1999), as a theory, aims to establish principles for conception, design, and implementation of computational Multi-Agent Systems. These systems of interacting agents are described in terms of components (agents), structure (network of agents), and organisation (ways and reasons for agents to interact).

As intentionality is embedded into the agents, they need mental-like processes for decision and action. Drawing from traditional psychology, AI tends to describe and explain human behaviour through mental states representing beliefs, desires, and intentions (Brazier et al. 2002). The Belief-Desire-Intention (BDI) paradigm, largely used in AI, states that individual decisions arise from the recursive exchange of information between these three mental states (Figure 3.1).

Figure 3.1. Belief-Desire-Intention (BDI) structure of a conative system



Source: Brazier et al. 2002

Jacques Ferber (1999: 242), in his design framework for MAS, proposes a more comprehensive classification of these mental states, he calls cognitons, for which Table 3.1 gives a partial list organised into categories. These different categories represent different sub-systems interacting during cognitive processes.

Table 3.1. Partial list of Cognitons proposed by Ferber

Category	*Cogniton*	*Description*
Interaction	Percept	Cogniton transmitted by external sensors
	Information	Cogniton transmitted by another agent
	Decision	Cogniton selecting action
	Request	Cogniton transmitted to another agent
	Norm	Cogniton imposed by the social organisation
Representation	Belief	Cogniton representing states of the world and self
	Assumption	Possible representation not yet believed
Conative	Tendency	Cogniton resulting from impulse or demand
	Impulse	Internal need coming from the conservative system
	Demand	External need resulting from request or percept
	Intention	Internal duty for decision
	Command	Cogniton selecting decision
	Engagement	External constraint on decision
Organisation	Method	Set of rules and techniques to implement action
	Task	Set of stages needed to implement action or method

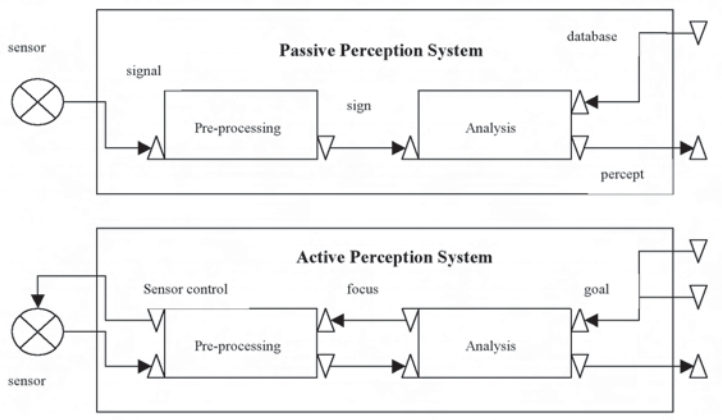
Source: Ferber 1999

Interaction system

The interaction system enables the agent to perceive and acquire information from the surrounding environment. This individual perception contributes to the elaboration of a subjective, limited, and contextual representation of the world.

From a philosophical perspective, there are two conflicting theories on perception. On one hand, the Aristotelian view assumes that perceived objects actively 'impregnate' our senses. Thus, we passively receive this imprint and integrate it to our cognitive system. This causal conception of perception asserts that we can access the objective qualities of surrounding objects. This model is widely accepted in cognitive science (logic theory) and computer science (shape recognition). On the other hand, the Kantian view asserts that percepts are constructed by the observer and depend upon previous experiences of perception. This active conception of perception constitutes an axiomatic principle in semiotics and it is consistent with major experimental results in neurobiology. Unfortunately, its application to computer science raises several technical problems that have, so far, limited its use in AI despite valuable experiments such as the 'Talking Heads' (Kaplan 2001). Figure 3.2 presents two computational systems of perception for artificial agents based on active or passive perception.

Figure 3.2. Passive and active perception systems. (Adapted from Ferber 1999)



System of representation

The system of representation enables the cognitive agent to store and manipulate acquired knowledge and beliefs. AI tends to group knowledge, know-how, experience, facts, and memories into a single set of information called 'beliefs'. These beliefs help the agent to decide and to implement actions. As a matter of fact, much theoretical work has been concerned with ways of representing, classifying and manipulating these beliefs for action (pragmatic dimension) rather than focusing on the very essence of 'knowledge' (epistemological dimension).

The Physical Symbol System Hypothesis, enunciated by Simon and Newell, is the founding principle of Symbolic Artificial Intelligence (SAI). Borrowing concepts from philosophy of mind (Kantian schemata) and semiotics (Peircean symbols), the principle states that any belief can be represented through a set of symbols and rules of inference (ontology). Four axiomatic propositions are generally accepted:

- Representations are independent from any underlying physical structure.
- Mental states are intentional: they are linked to a referent external to the agent.
- Representations are made of symbols or groups of symbols.
- Reasoning consists in manipulating symbols with rules of logic inference.

Evidence coming from neuro-biology has supported criticism of the first proposition by Connection Artificial Intelligence (CAI). The use of neural networks for task-oriented reasoning has indeed provided powerful alternate solutions. But, a more general criticism towards SAI relates to its implicit assumption of perfectly autonomous agents. As a matter of fact, social agents are embedded into an environment that not only supports and feeds individual reasoning but, more essentially, 'permeates' individual experiences through permanent interactions. Social psychology asserts that intelligence is culturally grounded and that knowledge evolves only through interaction with others by means of proposition, confrontation, and refutation (Cole and Scribner 1974). Interestingly, this same criticism gives some credit to the unique concept of belief used by SAI to describe different types of knowledge: from a social psychology viewpoint, any type of knowledge results from a historically contingent and socially built consensus. Hence, scientific knowledge and theories are themselves meta-beliefs, consensual models of a given 'reality'.

From an SAI perspective, agents continually use beliefs and assumptions (cognitons) from their system of representation to build descriptive or predictive models of their environment. Ferber (1999) proposes the following list of belief categories:

- Environmental belief: current or predicted state of the physical environment.
- Social belief: social norms and rules applicable within a given social group.
- Relational belief: competences and intentions attributed to other known agents.
- Personal belief: representation of self.

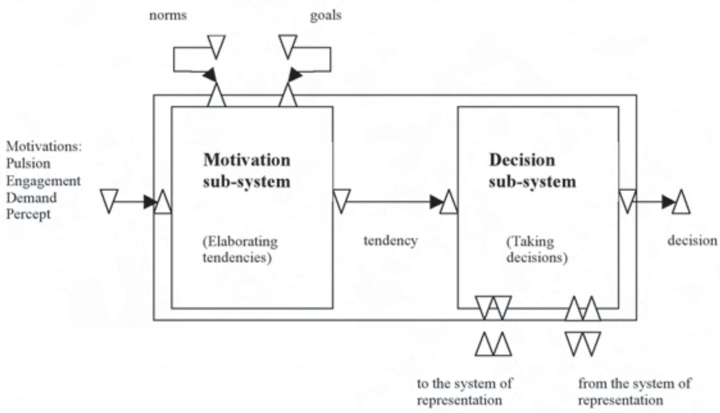
Conative system

The conative system defines the set of activities to be undertaken by an agent, based on available information and beliefs. The ways agents take their decisions, and the reason why they discard some options to focus on others, are questions that stretch well beyond Artificial Intelligence and nurture endless debates in philosophy and psychology.

As SAI relies upon logic inferences to describe an agent's behaviour, such as predicate logic or modal logic, causality links are meant to be rational. Hence, agents tend to display goal-satisfying decisions and, therefore, their actions are first driven by their needs and tendencies. According to Ferber (1999), these tendencies are themselves motivated by percepts, impulses, norms, or engagements, and trigger a decisional process based on existing beliefs and assumptions (Figure 3.3). Like beliefs, motivations can be separated into four categories:

- Environmental motivation: reflex or reinforcement due to percepts.
- Social motivation: engagement due to social norms or deontic rules.
- Relational motivation: engagement or hedonism linked to other interacting agents.
- Personal motivation: self-engagement or hedonism due to impulses.

Figure 3.3. Conative system and its two sub-systems. (Adapted from: Ferber, 1999)



The way SAI agents take intentional decisions and eventually undertake subsequent actions is largely based on causal philosophy of action (Bratman 1987). Intention is altogether a choice and an engagement towards this choice. Hence, in AI, an agent X is said to have the *intention* to perform an *action* A if X wants a *proposal* P about the state of the world to be true, and:

- X believes that P is a consequence of A,
- X believes that P is not currently realised,
- X believes he is able to perform A,
- X believes that A is possible and, consequently, P will be satisfied.

Cohen and Levesque (1990) have proposed a formalism for rational action, based on modal logic that has been largely used in Distributed Artificial Intelligence (DAI). Their formalism gives way to necessary, possible, or contingent predicates. Likewise, it allows expressing the temporality of intentions as *planning to do something in the future,* and needs a different set of tasks compared with *deciding to do something now*. Applying such formalism to Multi-Agent Systems implies that each agent is able not only to predict the consequences of its intended action, but also to anticipate the results of the other agent's behaviour. Hence, the agent needs to carry in his social or relational beliefs some ideas about the other agent's commitments. This is where the concept of engagement becomes paramount: self and social engagement are needed to introduce some sort of regularities in the system that can be hopefully anticipated.

Organisation system

Finally, the organisation system, through its methods and tasks (cognitons), allows the agent to prioritise, halt, and resume the pending decisions provided by the conative system. External information channeled through the interaction system may alter the implementation of a decision into action. A suspended decision will eventually resume according to the persistence of its triggering intention(s) (Ferber 1999).

Exploring human ecosystems

It is one thing to understand how SAI is used to design rational and intentional cognitive agents. But we also have to question the reasons why, in the first place, we intend to create these artificial entities? I will leave aside DAI applications belonging to robotics or computer-oriented technologies where autonomous agents tend to 'mimic' intentional cognition in order to perform actions considered as rational by their designers. After all, these agents are not supposed to be, or even to represent, human beings.

Instead, I will concentrate on these Multi-Agent-Based Simulations (MABS) that are meant to represent actual human ecosystems. Only a small proportion of those applications are used as social virtual experiments to explore cognitive processes. Relying on robust cognitive architectures inherited from SAI, these models are designed to help theoretical breakthroughs:

> I believe that the contribution of [Multi-Agent-Based Social Simulation] to the theoretical development of the cognitive and social sciences could be really remarkable. SS can provide not only an experimental method, but good operational models of cognitive 'actors', of the individual social mind, of group activity, etc. Models that can be richer, more various, and more adequate than those provided by economics, without being less formal. In particular, my focus on the core relation between functions and cognition was aimed at pointing out how the coming 'agent-based' approaches to social theory, using learning but deliberative agents, could deal with very old and hard problems of the social sciences and could re-orient them.
> (Castelfranchi 2001: 35)

As a matter of fact, a large proportion of MABS applications are designed to explore and understand complex interactions between actual actors and their environment. Bousquet and LePage (2004) or Hare and Deadman (2004) provide comprehensive reviews of these models. Most of these applications depart from the SAI paradigm and implement over-simplistic, task-oriented, rule-based agents, focusing on spatial interactions, social communication and individual mobility. Often, the drift from internally consistent and Formal Logic Compliant (FLC) agents is justified by synthetic information coming from field surveys or expert knowledge.

I would argue that in both cases, formal logic compliance or not, MABS will fail to deliver if we cannot find innovative ways to link the model, its object, and its interpreter. For deceptive idols and socially constructed icons are conspiring against agents.

Deceptive idols

Cognitive agents are supposed to behave rationally, as they represent rational human beings. FLC agents abide by a positivist and scientific rationality, and need a consistent set of decisions to act. Non-compliant agents (FLN) are usually designed according to the phenomenological interpretation of behaviour given by experts (i.e. sociologists, anthropologists). In both cases, the question is not about the acceptance of a rational behaviour. The question is about the axiomatic predicates used by the agents, or interpreted by the experts.

Bacon’s idols

In his *Novum Organum*, Francis Bacon classified the intellectual fallacies of his time under four headings which he called idols (Nova Organum 1620: 345:39):

> There are four classes of Idols which beset men's minds. To these for distinction's sake I have assigned names, calling the first class Idols of the Tribe; the second, Idols of the Den; the third, Idols of the Market Place; the fourth, Idols of the Theater.

An idol is an image, in this case held in the mind, which receives veneration but is without substance in itself. Bacon did not regard idols as symbols, but rather as fixations. In this respect he anticipated modern psychology.

'Idols of the Tribe' are deceptive beliefs inherent in the mind of man, and therefore belonging to the whole of the human race. They are abstractions in error arising from common tendencies to exaggeration, distortion, and disproportion. First of all, our tendency to let emotions rule reason can give us false impressions of the truth based on our feelings at the time. Another common idol lies in our tendency to seek out evidence of that which we already believe to be true. Bacon suggests that we become affectionate to ideas we have found and carried with us for some time; we become attached to them and collect evidence that supports them while throwing out that which contradicts them. Interestingly, Castelfranchi (2001) proposes a similar mechanism to explain the emergence of social functions among agents through ‘learning without understanding’ processes, superseding the intentional cognitive processes. But, so far, emotional agents remain out of reach of the current developments in SAI.

'Idols of the Den' (also called 'Idols of the Cave' in some editions) are those which arise within the mind of the individual. Like in Plato’s allegory, thoughts of the individual roam about in this dark cave and are variously modified by temperament, education, habit, environment, and accident. Thus an individual who dedicates his mind to some particular branch of learning becomes possessed by his own peculiar interest, and interprets all other learning according to the colors of his own devotion. In our case, this idol may affect directly the designer rather than the artificial agent. Lissack and Richardson (2001: 101) give an illustration of this bias in their criticism of Wolfram's claim that most complex systems can be accurately represented by rule-based atomistic models:

> The act of interpreting differs from the act of observing, and both may differ significantly from the underlying phenomenon being observed. In their failure to respect this distinction, strong MSM proponents [Wolfram and colleagues] are implicitly suggesting that the interpretation is reality. However, while a good model of complex systems can be extremely useful, it does not allow us to escape the moment of interpretation and decision.

Obviously, deterministic and limited expert knowledge used to design FLN agents suffers the same type of criticism. Causal rules of behaviour inferred by an expert are merely subjective interpretations of a given reality. Putting it simply, there are as many realities as there are experts. Agar (2005) recently proposed a way forward by means of coupled emic/etic approaches to social simulations. The author advocates a constant feedback between what makes sense for the actual actors depicted in the model (emic) and what seems meaningful to the designer (etic).

'Idols of the Marketplace' are errors arising from the false significance bestowed upon words, and in this classification Bacon anticipated the modern science of semantics. The constant impact of words variously used without attention to their true meaning often betrays their purpose, obscuring the very thoughts they are designed to express. Acknowledging the volatility of the 'true meaning'of words, Bacon just caught a glimpse of the active perception, theorised by Kant and Peirce later on. Words, as elementary percepts, do not carry any specific and intrinsic meaning when they are perceived. They have to be re-interpreted internally by the receiver, according to previous knowledge and environmental hints. Maturana and Varela (1980: 32) propose to drop the denotative understanding of language altogether in favour of a connotative approach:

> So long as language is considered to be denotative it will be necessary to look at it as a means for the transmission of information, as if something were transmitted from organism to organism, in a manner such that the domain of uncertainties of the 'receiver' should be reduced according to the specifications of the 'sender'. However, when it is recognized that language is connotative and not denotative, and that its function is to orient the orientee within his cognitive domain without regard for the

> cognitive domain of the orienter, it becomes apparent that there is no transmission of information through language. It behooves the orientee, as a result of an independent internal operation upon his own state, to choose where to orient his cognitive domain; the choice is caused by the 'message', but the orientation thus produced is independent of what the 'message' represents for the orienter. In a strict sense then, there is no transfer of thought from the speaker to his interlocutor; the listener creates information by reducing his uncertainty through his interactions in his cognitive domain.

As stated earlier, mainstream SAI satisfies itself with passive perception processes. Hence, most FLC agents can receive intelligible and meaningful information from other agents, without having to engage into deciphering and re-interpretation stages. Somehow, it makes agents' lives seem much easier than ours!

'Idols of the Theatre' occur due to sophistry and false learning. These idols are built up in the fields of theology, philosophy, and science, and, because they are defended by learned groups, are accepted without question by the masses. When theories have been cultivated and have reached a sufficient level of consensus they are no longer questioned. The long standing hegemony of the symbolico-cognitivist paradigm cannot hide the fact that relevant alternatives have challenged SAI's dominion: connectionist and evolutionary theories perform better on learning processes (Kaplan 2001); the autopoietic theory, by refusing the conveyance of information through linguistic interaction, provides a unified and unchallenged approach to signaling interactions (verbal, non-verbal, or extra-verbal) through structural coupling between individuals (Maturana and Varela 1980).

Prospect theory

Prospect theory focuses on cognitive and psychological factors that determine the value of risky prospects (Kahnemann and Tversky 2000). Its initial assumption is that subjective values attached to gambling are carried by expected changes of wealth (gains or losses) rather than ultimate states of wealth. More importantly, prospect theory replaces the traditional concept of risk aversion by a more intuitive one, called loss aversion, by which people tend to consider that a loss of $X is more averse than a gain of $X is attractive. This assumption explains why people might be risk seeking, and no longer risk averse, in the domain of losses.

Though trivial from an espistemological viewpoint, Prospect theory tends to reconcile theoretical development with empirical facts. According to Kahnmann and Tversky (2000:1):

> The study of decision addresses both normative and descriptive questions. The normative analysis is concerned with the nature of

> rationality and the logic of the decision making. The descriptive analysis, in contrast, is concerned with people's beliefs and preferences as they are, not as they should be.

Although based on formal logic predicates, Prospect theory recognises the fact that part of the knowledge necessary to complete the theory is inherently elusive. It is possible to design FLC agents founding their decisions on this theory, but the axiomatic predicates that will tell us about discrepancies between gains and losses, or the way different items will be affected by these discrepancies remain, by far, out of reach.

Furthermore, prospect theory threatens directly two logical pillars of decision theory traditionally used by SAI: preference invariance and value coincidence. Invariance requires that the preference order between prospects should not depend on the manner in which they are described. Kahneman and Tversky (2000:5) have demonstrated that invariance cannot generally be satisfied: 'invariance is normatively essential, intuitively compelling, and psychologically unfeasible' due to the framing of outcomes through formulation effects. The framing effect also affects the relation between experience and decision values. But rational agents seldom make a difference between experience values (direct outcomes of actual actions) and decision values (expected outcomes of an anticipated choice). These two values tacitly coincide, despite Kahnemann's and Tversky's (2000: 16) warning:

> Some factors that affect experience are not easily anticipated, and some factors that affect decisions do not have a comparable impact on the experience of outcomes.

Cognitive dissonance

For Bacon, knowledge is intimately mixed with the idols, hence prefiguring our modern concept of belief. More importantly, Bacon draws visionary consequences from the presence of the idols, in terms of communication (Nova Organum 1620: 346:35):

> enter quietly into the minds that are fit and capable of receiving it; for confutations cannot be employed, when the difference is upon first principles and very notions and even upon forms of demonstration.

Individuals and groups exhibit varied responses when faced with new information. If such information is consistent with extant behaviours and beliefs, it can be readily accepted and integrated. However, if the new information conflicts with behaviour and belief, the resulting state is described as 'cognitive dissonance' (Bradshaw and Borchers 2000). According to the theory, the inconsistency and psychological discomfort of cognitive dissonance can be reduced by changing one's beliefs, values, or behaviour. Dissonance can also be avoided by rejecting or avoiding information that challenges belief systems or by interpreting dissonant information in a biased way. In this regard, most SAI structures force agents to discard new information conflicting with a given set of consistent predicates. It is only through reinforcement (punishment or reward) due to experience that contrasted set of predicates can be established. Elaborating on the conflicting views upon 'uncertainty' between scientists and policy-makers to explain the science-policy gap, Bradshaw and Borchers (2000: 3) outline the complexity of cognitive dissonance:

> Dissonance between existing beliefs and new information may be shaped by a host of factors, all of which inhibit the rate at which scientific findings are assimilated into policy. In what we have called the 'volition' phase of the science-policy gap, public debate around an emerging scientific consensus may derive from a combination of cultural, psychological, and economic interests threatened by the policy inferences of dissonant scientific findings.

The authors particularly point at the contrasted rhetorical figures used by scientists when they are in charge of policy-making compared with their usual handling of scientific uncertainties. Designers in DAI have tried to encapsulate these internal cognitive conflicts by implementing Agent-Group-Role structures in which one agent belongs simultaneously to several socio-cultural groups and plays different roles accordingly (Ferber, 1999). But the internal conflict resolution mechanisms—for example, between friendship engagement and professional commitment—rely on formal and individualistic logic once more.

Overall, despite the increasing interest in, and use of Multi Agent Systems to represent human ecosystems, the SAI paradigm remains mostly unchallenged as a theoretical framework used to develop cognitive agents. But looking at the real world through the lenses of the social sciences, we have to acknowledge the fact that a meta-theory of human behaviour doesn't exist yet. Rational decision theory, prospect theory, social learning theory, and others give us partial clues about human behaviour, and none can stand as an overarching and unified framework.

Hence, we must handle cautiously Cederman's assertion about the 'New Deal' offered by Multi Agent Systems to social scientists. These tools and their current states of application generally rely upon reductionist views of the world: symbolico-cognitive hypothesis, formal predicate logic, rational decision, and autonomy. As a matter of fact, the SAI paradigm stands out as a nomothetic meta-model for agent's behaviour while its foundation doesn't represent a meta-model of human behaviour but merely a partial interpretation of it.

Socially constructed icons

Beside the symbolico-cognitive paradigm, Multi Agent Systems explicitly emphasise the autonomy of agents through physical integrity, autonomous decision, and rational action. But social psychology (Cole and Scribner 1974), psychology of development (Papert 1980) and collective action theory (Oliver 1993) insist, for different reasons, on the fact that human beings are social beings before anything else: learning processes, cognitive inferences, or individual actions are shaped and dictated by our environment. Recent developments in DAI have tried to incorporate this 'sociality' dimension into agent's behaviour. We'll come back in the next section to Hogg's and Jennings' (2001) proposal to include social rationality into agents' expected utilities, or Jager's and Janssen's (2003) attempt to consider basic human needs and uncertainty as the driving factors behind decision making processes of their agents. But first, we need to understand the elements of social cognition that weaken the foundations of the SAI paradigm.

Peirce's icons

Charles Sanders Peirce, a founder of modern semiotics, has asserted that:

- we cannot think without signs,
- we have no ability for intuition, all knowledge flows from the former knowledge,
- we have no ability for introspection; all knowledge about the inner world is produced by hypothetical reasoning on the basis of observation of outer things. (Peirce Edition Project 1998)

On this solid foundation he builds up his entire theory of signs. Being a pure positivist himself, at least during his early career, Peirce is convinced to build a theory based on formal logic, independent from particular minds, and characterised by a triadic relation (Peirce Edition Project 1998: 411):

> [semiosis is an] action, or influence, which is, or involves, a cooperation of three subjects, such as a sign, its object, and its interpretant, this tri-relative influence not being in any way resolvable into actions between pairs.

Peirce suggests that people have not and cannot have direct access to reality. Signs are nothing else than the universal medium between human minds and the world. Semiotics defines different categories of signs according to their level of abstraction, from icons to symbols (Peirce Edition Project 1998). As a primary and elementary sign, an icon looks like it is signified. There is no real connection between an object and an icon of it other than the likeness, so the mind is required to see the similarity and associate the two. A characteristic of the icon is that by observing it, we can derive information about its significance. Hence, semiotics embrace the Kantian concept of active perception as stated previously.

Since signs are not private but socially shared, it is society that establishes their meaning. Therefore, the transcendental principle in semiotics is not (divine) intuition, but community, and the criterion of truth, social consensus. According to Peirce, any truth is provisional, and the truth of any proposition cannot be certain but only probable. Umberto Eco, in his exhilarating *Kant and the Platypus*, provides a vivid description of this fallibilism principle (Eco 1997: 98):

> Naturally at this point transcendentalism will also undergo its Copernican revolution. The guaranty that our hypotheses are 'right' will no longer be sought for in the a priori of the pure intellect but in the historic, progressive, and temporal consensus of the Community. Faced with the risk of fallibilism, the transcendental is also historicized; it becomes an accumulation of interpretations that are accepted after a process of discussion, selection, and repudiation. This foundation is unstable, based on the pseudo-transcendental of the Community.

In order to sort out conceptual confusions born from consensual uncertainties, Pierce uses a 'pragmatic' approach by linking the meaning of concepts to their operational or practical consequences. This pragmatism will act later on as a corner stone for SAI's intentionality of beliefs. But, when conventional SAI says: 'beliefs *are* intentional', semiotics tell us: 'we *have to consider* beliefs as intentional'. Somehow, SAI has developed an hyper-positivist approach of cognition, through passive perception, intentionality, and formal modal logic that has gone beyond the scope of its theoretical background. In this regard, FLN agents are more likely to 'mimic' social behaviour compared with their FLC counterparts as the rationale for their actions is inferred from an holistic, though subjective, view of the social system constructed by their designer. In other words, some behavioural rules have to be considered for the FLN agents to interact in a consistent way (system level), without prerequisite conditions on cognitive consistency (agent level). But we'll see in the next section that it comes at a cost.

Enactive cognitive theory

We human beings are living systems that exhibit cognition, and there is no way for us to address, much less explain, our cognitive abilities without employing those same cognitive abilities. To date, the primary response to this paradox has been to ignore it and proceed with respect to a presumably fixed fundament, external to our act(s) of cognition. Where the presumptive fundament is an 'objective reality', the mediation between situation and action is explained in terms of ordered inference with respect to a model of that reality.

Maturana and Varela (1980) questioned this conventional approach when it is confronted with the tacit, extralinguistic, or emotive character of human behaviour. They further disputed the 'objective reality' concept by considering:

- Ourselves operating in multiple 'worlds', particularly socio-cultural ones.
- A 'world' being molded by contextual factors intertwined with the very act of engaging it.

Their autopoietic theory considers living beings as living systems embedded into larger systems constituted by themselves and the environment they interact with. The theory addresses the basic configural and operational circularities of these living systems (Maturana and Varela 1980). Unlike other more positivist approaches to complex systems, the autopoietic theory focuses on the observer himself. It accomplishes this by shifting explanatory focus from atomic units in an objective world to essential relations among processes operating in circular ways to constitute the organism as a living system and the observer as a cognitive organism. As clearly stated by Maturana (Maturana and Varela 1980: 7):

> Cognition is a biological phenomenon and can only be understood as such; any epistemological insight into the domain of knowledge requires this understanding.

Later on, Varela and colleagues brought phenomenological concerns into the world of cognitive science (Varela et al. 1991). Their goal was to incorporate everyday experience into the scope of studies which had heretofore addressed cognition in terms of disembodied rational processes, circumscribed by abstract beliefs purported to mirror an objective milieu. All concepts fully accepted by symbolico-cognitivist approaches. Varela and colleagues proceed from the assumption that experience necessarily predates and underpins enquiry. Maintaining a focus on experience as action allows inspection and reflection on the manner in which 'mind' and 'body' reciprocally engage to consummate experience. Midway between cognitivist and connectionist paradigms, their Enactive Cognitive theory considers that (Varela et al. 1991: 148):

> context-dependent know-how [shouldn't be treated] as a residual artifact that can be progressively eliminated by the discovery of more sophisticated rules but as, in fact, the very essence of creative cognition…Knowledge depends on being in a world that is inseparable from our bodies, our language, and our social history—in short, from our embodiment.

As such, enactive cognitive theory does not address cognition in the currently conventional sense as an internal manipulation of extrinsic information or signals, as semiotics or symbolico-cognitivism would have us believe (we have seen above how Maturana and Varela treat linguistic interactions) . Instead, it grounds cognitive activity in the embodiment of the actor and the specific context of activity. The theory fits very well with current trends toward emphasising contextualised studies of humans, their interactions, and their social systems. Unlike symbolico-cognitivist approaches, the enactive cognitive theory is consistently ‘relativistic’ in the sense that any given observation is observer-bounded, history-contingent, and socially embedded.

The theory implies an epistemology analogous to that of constructivism (Funtowicz and Ravetz 1993). For this reason, Multi Agent Systems built from FLN agents and based on expert knowledge never describe *the reality as it is* but rather *a probable reality* interpreted by a given observer in a given context. It means that such models cannot be evaluated through traditional scientific methods but need to be assessed, in context, against a set of criteria intelligible to the observed subjects and to the observer.

Social constructions

As mentioned above, very influential studies in social sciences have focused on the embodiment of cognitive activities into the actor’s experiences. In the psychology of development, Piaget asserted that children only develop through progressive interactions with their environment (Papert 1980). In the early stages of development, Piaget’s theory suggests that learning comes *only* from unexpected changes in the environment of the subject. Changes create a cognitive imbalance that must be adjusted through assimilation, the new experience is integrated into the current view of the world, or accommodation, the current view of the world must be modified to fit in the new experience. From this constructivist viewpoint, imbalance is no longer considered as a negative factor but rather as a driving force towards cognitive development. Hence, a perceived environmental unstability seems to trigger some sort of phase transition, in a complexity theory sense, from one state of representation to another. It is interesting here to draw a link with the enactive cognitive theory, for which concrete experiences (know-how) are driving cognitive processes through connotative interactions (Maturana and Varela 1980).

Indeed, a new picture of cognitive processes takes shape; a mind submitted to perpetual changes and characterised by a dynamic equilibrium between past and current views of the world; a mind dominated by inductive inferences and desperately trying to make sense of incoming information through deductive and formal logic (Batten 2000). This description sharply contrasts with the well structured and perfectly organised mind proposed by SAI designers. As a matter of fact, FLC agents process information in a consistent way; their actions are driven by intentional cognition and the experienced outcomes are evaluated against expected utilities (Brazier et al. 2002).

Marvin Minsky, in his famous *Society of Mind* (1985), proposed an alternate solution to the symbolico-cognitivist theory. Drawing from an initial intuition that our mind is made of structural and independent units called *frames*, Minsky views the human mind itself as a vast society of individually simple processing *agents*. The agents are the fundamental thinking entities from which minds are built, and together produce the many abilities we attribute to minds. The advantage in viewing a mind as a society of agents, as opposed to the consequence of some formal logic-based system, is that different mental agents can be based on different types of processes with different purposes, ways of representing knowledge, and methods for producing results. The consistency of this atomistic mental system rises from the organisation of interactions between agents, not from a unified and coherent reasoning (Minsky 1985: 308):

> What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.

In principle, these networks of mental agents display dynamic interactions where agents can be modified, created, or deleted in order to adapt or to respond to a given experience. This interactionist theory can accommodate—though not explicitly formulated—constructivist (Papert 1980) and connotative (Maturana and Varela 1980) standpoints. Unfortunately, to date, few Multi Agent Systems have been implemented according to Minsky's views, and a vast majority of these models concern robotics or computer-oriented technologies. Both fields have been excluded from the scope of this paper.

Another advantage of the interactionist theory is that it creates a continuum between individual and social actions. The same paradigm is used from mental processes to collective action, from personal engagement to societal norm enforcement. As a matter of fact, sociologists studying social movements often refer to Collective Action Frames to represent the process by which activists create and modify action-oriented sets of beliefs and meanings that inspire and legitimate the activities of a social movement (Benford and Snow 2000). A crucial aspect of collective action frames is that they are not merely aggregations of individual attitudes and perceptions but also the outcome of negotiating shared meaning. Collective framing necessitates complex interactions between activists, antagonists, bystanders, and observers (op. cit.: 614):

> Frames help to render events or occurrences meaningful and thereby function to organize experience and guide action. Collective action frames also perform this interpretive function by simplifying and condensing aspects of the 'world out there', but in ways that are intended to mobilize potential adherents and constituents, to garner bystander support, and to demobilize antagonists.

Cederman's quest for generative explanations of social forms based on configurative ontologies cannot be entirely satisfied by autonomous and rational agents, as formal modal logic constitutes a very specific and limited type of ontology. As we have seen, the interactionist paradigm offers a much richer context for cognition and *sociation*. Unfortunately, technical caveats have prevented so far to implement purely interactionist computer models beyond experimental and limited applications. Finally, Cederman's call to the Simmelian tradition questions the nature of these social patterns we are supposed to observe: if we accept Maturana's and Varela's views, these patterns are strongly observer-dependent and they need a constructivist epistemology for authentication.

Paradigm shift

From the previous sections, we can infer the following:

- The symbolico-cognitivist paradigm relies upon a limited, fragmented, and sometimes conflicting understanding of cognition and behaviour.
- Autonomy of cognition is inherently limited as beliefs are socially constructed and experience preempts cognition.
- A model of social behaviour – as a tentative representation of an objective reality – cannot be separated from its designer's experience and viewpoint.
- Multi Agent Systems are specific models that display autonomous and cognitive entities, hence, they are concerned with the three previous limitations altogether.

Systems built with FLC agents suffer mainly from the first and second limitations. They tend to limit the subjectivity of the representation by adhering to a consensual theoretical framework of reference. But the SAI paradigm itself could be considered, in a Baconian sense, as an idol of the theatre altogether. Systems built with FLN agents suffer mainly from the second and third limitations. They accept our limited understanding of cognition by way of pragmatism. But they are intimately related to their designer's experience.

In all cases, the power of explanation of a given model can hardly be assessed through traditional scientific positivism as soon as human cognitive processes are simulated. If it were the case, a set of explanatory hypotheses could be used to unambiguously validate the simulated processes against objective criteria. So far, in the absence of any meta-theory of human behaviour, only experimental economics and experimental psychology are used to inductively validate these models, with all the unrealistic constraints imposed to these experiments. Nevertheless, our next section provides meaningful examples of attempts to refine the symbolico-cognitivist approach in order to match the reality of socially embedded behaviours.

As mentioned above, Agar (2005), dealing with very elusive behaviours of illicit drug users, advocates a dual approach of modelling where actor-based components (emic) are used along with theory-based components (etic) in order to implement realistic behavioural and social models. Of course, this proposal assumes that ethnography or sociology can provide *in situ* replicable and explicit methods of knowledge elicitation. Dray et al. (2006) have recently tried to formalise such a process, using knowledge engineering techniques. If these techniques allow, in principle, a better traceability of elicited knowledge we have to consider Cole's and Scribner's (1974) findings from their cross-cultural studies: any given ethnographic experimental setting reflects its designer's own beliefs and intentions. Therefore, the authors advocate a constant cross-validation between the observer and the observed subjects in order to limit misinterpretation. Maturana and Varela (1980) would label these iterative and circular interactions between a designer, an observer, and observed subjects, a structural coupling process. It is the fundament of the post-normal approach to modelling proposed by Funtowicz and Ravetz (1994) and described below.

Normal views on social rationality

Kluver and colleagues (2003) point out the main difference in building theories between social sciences and natural sciences. The latter always proceed through increment, starting with rather simple models and enlarging them by successive steps, whenever the progress of research made it necessary. This *normal science* is guided by a certain paradigm that is transformed into an increasingly complex model as long as the paradigm makes it possible. By contrast, social sciences often adopt an all-at-once approach by which theorists try to capture from the beginning as much of social complexity as they can within their conceptual framework. Kluver and colleagues have no doubt about the fact that social sciences must embrace a *normal* posture in order to progress towards a better understanding of social complexity (2003: 3):

> That is not possible in computational or mathematical sociology, respectively: the basic models must be simple in order to understand their behaviour in principle. The enlargement of the basic models that is always necessary in advancing research can only be done if this basic understanding has been achieved. Therefore the social sciences have to adopt this methodical procedure from the natural sciences if a formally precise theory of social complexity is to be achieved.

The SAI paradigm has supported, through normal and positivist procedures, the implementation of always more complex models of individual decision and social behaviour. If emotional decisions or the 'embodiment' of cognition are still on the shelves, the social rationality of agents, as opposed to their goal-satisfying rationality, has recently attracted much attention. Hogg and Jennings

(2001: 382), recognising that individual and social concerns often conflict leading to the possibility of inefficient performances, proposed a framework for making socially acceptable decisions, based on social welfare functions, that combine social and individual perspectives:

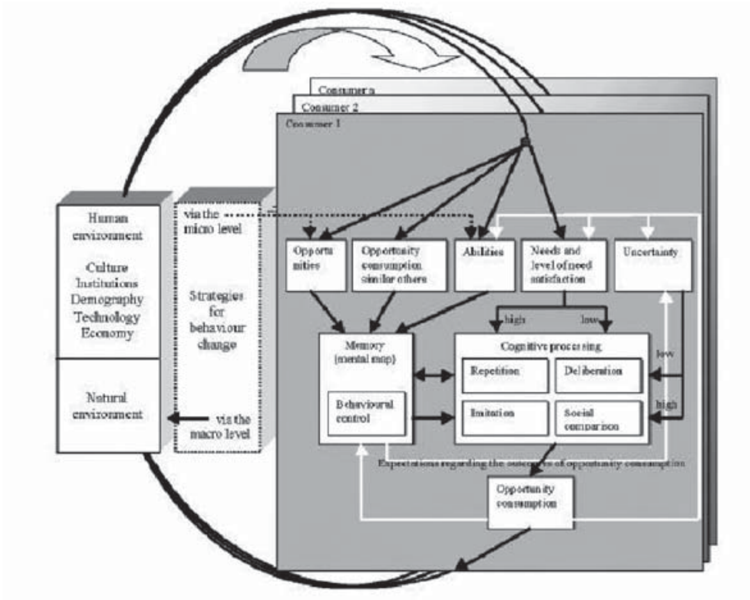
> To be socially rational, an individual maximizes his social welfare function over the different alternatives. This function represents how an individual agent may judge states of the world from a moral or social perspective by taking into consideration the benefit to others of its course of action and weighing it against its own benefits.

The dynamic balance between individual and social utility functions is controlled by a metalevel controller that considers available resources (adaptation) and past experiences (learning) to tune the amount of cognitive efforts put into social rationality. An equivalent mechanism is used by Castelfranchi (2001) in order to formalise social functions (or roles) assumed by individuals as non-intentional mental processes. In this case, a metalevel controller supersedes the intentional and rational system of the agent. A learning-without-understanding process reinforces mechanically some individual beliefs, whether they are beneficial for the agent or not.

In both cases, the cognitive architecture becomes increasingly complex and the *normality paradigm* appears to act as a 'patching' process applied on a system of formal inferences that was not designed for such a purpose in the first place. Furthermore, these architectures and organisations seldom rely on direct evidence for validation. Instead, *virtual social experiments* are used to generate results that are evaluated against plausible utility values at the system and individual levels.

In order to overcome the increasing complexity of SAI architectures, and to facilitate, to some extend, the direct validation of the building assumptions, Jager and Janssen (2003) propose an alternate formalism. Their consumat theory is to be considered one of these conceptual jumps characteristic of the evolution of *normal* science: a new paradigm supports the creation of simpler models compared with the previous ageing generation. The consumat approach considers basic human needs and environmental uncertainties as the driving factors behind decision making processes. Agents engage in different cognitive processes, including social imitation and comparison, according to their perceptions of individual needs and environmental threat (Figure 3.4). Hence, the consumat paradigm overrides two structural limitations of the symbolico-cognitivist paradigm: experience preempts cognition and social rationality is directly built into intentional processes. Being at an early stage of development, the new theory relies on relatively simple rules that can be validated against experimental economics settings.

Figure 3.4. Structure of a Consumat agent



Source: Jager and Jannsen 2003

A post-normal temptation

Funtowicz and Ravetz (1993) also use environmental uncertainties, along with decision stakes, to analyse problem-solving strategies in the context of environmental and population risk policy issues. They argue that traditional scientific methodologies are ineffective when either attribute is high. Instead, they propose a new scientific posture they call *post-normal science* (op. cit.: 739):

> In those circumstances, the quality assurance of scientific inputs to the policy process requires an 'extended peer community', consisting of all those with a stake in the dialogue on the issue. Post-normal science can provide a path to the democratization of science, and also a response to the current tendencies to post-modernity.

The dynamic of resolution of policy issues in post-normal science involves the inclusion of an adequate set of legitimate participants in the process of quality assurance of the scientific inputs. For example, persons directly affected by an environmental problem will have a keener awareness of its symptoms. Thus, they perform a function analogous to that of peer-reviewers in traditional science, which otherwise might not occur in these specific contexts. Closer to the concern of this chapter, Funtowicz and Ravetz (1993: 745) challenge the commonly admitted rationality of decision and action:

> Until now, with the dominance of applied science, the rationality of reductionist natural-scientific research has been taken as a model for the rationality of intellectual and social activity in general. However successful it has been in the past, the recognition of the policy issues of risk and the environment shows that this ideal of rationality is no longer universally appropriate. The activity of science now encompasses the management of irreducible uncertainties in knowledge and in ethics, and the recognition of different legitimate perspectives and ways of knowing.

Now, let's put this *post-normal* scientific posture into the context of our chapter. Among the majority of MABS used to explore human ecosystems, a significant number are meant to demonstrate how individuals and populations interact with their environment, as well as the environmental and social consequences of these interactions (Bousquet and LePage 2004). These situations are often characterised by:

- The presence of different groups of actors with contrasted, even conflicting, strategies.
- Irreducible uncertainties in representing and predicting responses from the environment.
- Individual and social rationalities based on multiple and competing utility functions.
- Self-referential conditions limiting goal-satisfying decisions to sub-optimal solutions.
- Emotional and cultural responses to policy incentives or penalties.
- Important framing effects and asymmetry of information.

We have to accept the fact that Multi Agent Models, even the more sophisticated ones, will always be pale copies of the original, subjective and partial representations of a dynamic reality. But recognising this very peculiar fact doesn't mean that these models are useless, even Lissak and Richardson (2001: 105) in their criticism of computer-based social models admit that:

> There is no need for the models in question to have predictive power, despite the strong desire of both consultants and their clients that such models 'work'. The pedagogical value of exploring the interactions of complex relations through the manipulation of models is more than enough to justify the efforts that go into model development and proliferation. Clearly, it is easier to manipulate a computer model than a fully fledged 'in reality' laboratory experiment, but the limitations of such models must be remembered.

Chefs-d'Oeuvres are chiselled by talented craftsmen, not by their tools. Talent is all about sharing a vision and choosing the appropriate tools. In this regard, we must admit that the ultimate criteria of validation for this type of model is its actual appropriation by the final users (policy-makers, local communities, or else) and the consensual acceptation of the simulated outcomes. In this case, we can infer that:

Due to irreducible uncertainties and complex interactions within human ecosystems, social simulation designers should abide by a principle of subsidiary formalism. The principle acknowledges the fact that the most reliable source of knowledge about human decisions and behaviour are indeed the actors of the real drama themselves. Hence, the principle of subsidiarity stipulates that designer's cognitive efforts should be directed towards engaging with real stakeholders and eliciting their own mental models in the first place. Only when this option is unrealistic, the designers shall make it clear to everyone that decisional rules and algorithms are derived from theoretical or empirical predicates.

The subsidiarity principle offers a non-threatening opportunity for ethnography to accept the challenge of a new and dynamic formalism, beside narratives (Lansing 2003). The principle also mitigates the accusation of mere indexicality uttered by *System Thinkers* against individual-based simulations (Lissak and Richardson 2001). Finally, a post-normal posture might help solve the problem of validation that any complex system model is faced with: partial scientific validation associated with social authentication legitimate the modelling process and its outcomes (Funtowicz and Ravetz 1993).

Companion modelling

During the early 1990s, Granath (1991) introduced the concept of *collective design* in industry to define a process by which all actors involved in the production, diffusion, or consumption of a product are considered as equal experts and invited to participate to the design of the product. Each expert actively contributes to the collective process of transdisciplinary creation. Collective design usually faces two problems:

- socio-cultural barriers between different disciplines or social groups; and
- heterogeneous levels of knowledge and dissonant modes of communication.

Hence, in order to implement a collective design process, it is important to initially elicit specific knowledge and practices among experts. Then, the process itself must be grounded into successive *mediating objects*. These artifacts (ideas, models, products) are meant to channel creativity and to structure communication among experts. Collective design is to be considered as a social construct, no longer functional or rational. The final product emerges from conflicts, alliances, and negotiations.

During the late 90s, collective design and other approaches like system thinking or action learning were appropriated and adapted by scientists working on natural resource management (Hagmann et al. 2002; Barreteau 2003; D'Aquino et al. 2003). Interestingly, most of the cases were characterised by:

- direct engagement of science into field management issues (R&D projects);
- complex and adaptive socio-ecological systems (mid-scale with recursive interactions); and
- important cross-cultural contexts (rural Africa or Asia).

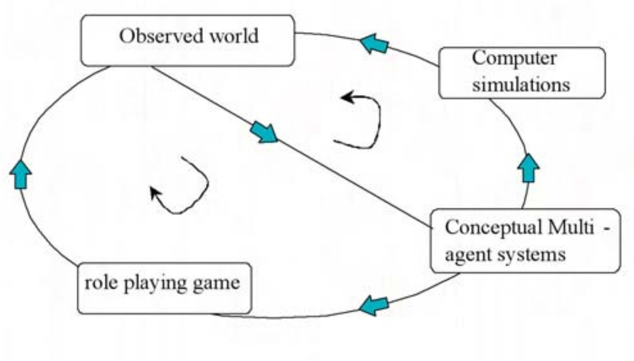
The co-construction of these models with local stakeholders didn't intend to provide normative models of reality, instead they were meant to enhance discussion and collective action through interactions around and about the mediating object (Lynam et al. 2002: 2):

> It is important to emphasize that, in the contexts in which these case studies are presented, the models were used more as part of a process of developing and exploring a common understanding of problems and possible solutions. They were not designed to be highly validated, predictive models in the sense in which systemic models are usually developed and used. We are not aware of other examples in which local people, who have no history of computer-based modeling, have been involved, not only in the use of computer models, but also in their development.

In these models, agents are designed according to the consensual information provided by their real counter-parts and the people they interact with. Decisional or behavioural rules are as complex and rational as the creators wish them to be. Likewise, agent's beliefs are tailored according to the phenomenological expression of the real stakeholder's mental models (Dray et al. 2006). Hence, this constructivist and post-normal modelling of cognitive processes doesn't intend to tell *How does it work?*, but rather *What is there that is so important?* But Becu and colleagues (2003) give evidence that the phenomenological expression of personal beliefs and intentions might not provide a consistent enough set of rules or assumptions to be directly encapsulated into the agent.

Recently, a group of scientists, *Collectif ComMod*, has decided to formalise their approach in order to establish deontic rules for developing companion models by direct interactions with local stakeholders (Collectif ComMod 2005). Companion modelling (ComMod) is an approach making use of simulation models in a participatory way to understand and facilitate the collective decision-making process of stakeholders sharing a common resource. The principle is to identify the various viewpoints and subjective referents used by the different stakeholders, and to integrate this knowledge into simulation models that serve as mediating objects. There is an iterative process of confrontation between factual evidence, model design, and scenario exploration (Figure 3.5). Other mediating objects or methods are usually used in conjunction with computer simulations, like role-playing games (Barreteau 2003; Dray et al. 2006). The different stakeholders, including researchers, aim at working out a common vision of the common resource management that highlights the diversity of interests. This approach has attracted growing attention from decision-makers and community-based organisations in order to engage more dynamically with stakeholders.

Figure 3.5. Traditional cycle of interactions during a companion modelling process



Conclusion

We have taken Cederman's assertion as a pretext for a journey through the world of artificial agents. From SAI to ComMod, we have drifted from positivism to constructivism, from a *normal* kind of science to a *post-normal* one. Despite the weaknesses and flaws pointed at in this chapter, the author, like Cederman, believes that Multi Agent Systems (MAS) offer a fantastic opportunity for the social, natural, and computer sciences to come together and engage in a 'new kind of science', much more challenging in epistemological terms than Wolfram's original proposal (see Lissak and Richardson 2001). Unlike Kluver and colleagues (2003), we think that there is a way for social sciences to appropriate computer formalisms without falling into a reductionist and positivist stand. The enactive cognitive theory (Maturana and Varela 1980) and the interactionist theory (Minsky 1985) need to be re-visited in a transdisciplinary manner. They can provide a theoretical substance to a vast majority of MAS applications built with non-formal logic compliant agents. The validation of these models needs to be embedded into a constructivist perspective where designers, users, and stakeholders not only evaluate the simulated outcomes, but also participate in the modelling process itself (Granath 1991; Collectif ComMod 2005).

Progressing on to the constructivist path doesn't mean that we have to discard the symbolico-cognitivist paradigm altogether. The strong, but limited, edifice proposed by SAI has shown undisputable capacities in producing replicable methods to test assumptions or benchmark findings on rational decision and action. For example, recent work from Castelfranchi (2001) or Hogg and Jennings (2001) are coming closer to designing realistic socially rational agents. But, paraphrasing Lissak and Richardson, 'the limitations of such models must be remembered'. Beyond the scope of this paper, the same type of warning must be addressed to social network theory and its extensive use of graph theory to explain social interactions (Borgatti and Foster 2003). Graphs are merely symbols, in a Peircean sense, displaying some sort of likeness with reality, they are not reality.

Finally, we have to agree with Cederman (2005) on the potential usefulness of Multi Agent Systems for social scientists. But we have to be aware of the fact that the symbolico-cognitive paradigm inherently limits the range of generative ontologies to be created. We have to take even more cautiously his reference to Simmelian social patterns. These patterns are subjectively construed by the observer, they are not the objective reality. Using MABS to replicate these patterns imposes conditions on the artificial agents that need to be socially validated as far as traditional scientific validation is no longer relevant in the case of complex human ecosystems.

Acknowledgement

The author wishes to thank David Batten, Francois Bousquet, and Roger Bradbury for their useful comments and advice.

References

Agar, M. (2005) Agents in living color: towards emic agent-based models. *Journal of Artificial Societies and Social Simulation* 8: 1. Available at http://jasss.soc.surrey.ac.uk/8/1/4.html

Bacon, F. (1620) *Novum Organum*. Trans. by Basil Montague. *The Works*, Vol. 3, pp. 343–71. Philadelphia: Parry and MacMillan (1854 edn). Available at http://history.hanover.edu/texts/Bacon/novorg.html

Barreteau, O. (2003) The joint use of role-playing games and models regarding negotiation processes: characterization of associations. *Journal of Artificial Societies and Social Simulation* 6(2). Available at http://jasss.soc.surrey.ac.uk/6/2/3.html

Batten, D.F. (2000) *Discovering Artificial Economics. How Agents Learn and Economies Evolve*. Oxford: Westview Press.

Becu, N., F. Bousquet, O. Barreteau, P. Perez and A. Walker (2003) A methodology for eliciting and modelling stakeholders' representations with Agent Based Modelling. *Lecture Notes in Artificial Intelligence* 2927: 131–49.

Benford, R.D. and D.A. Snow (2000) Framing processes and social movements: an overview and assessment. *Annual Review of Sociology* 26: 611–39.

Borgatti, S.P. and P.C. Foster (2003) The network paradigm in organizational research: a review and typology. *Journal of Management* 29(6): 991–1013.

Bousquet, F. and C. Le Page (2004) Multi-agent simulations and ecosystem management: a review. *Ecological Modelling* 176(3-4): 313–32.

Bradshaw, G.A. and J.G. Borchers (2000) Uncertainty as information: narrowing the science-policy gap. *Conservation Ecology* 4(1): 7. Available at http://www.consecol.org/vol4/iss1/art7/

Bratman, M. (1987) *Intentions, Plans, and Practical Reason*. Harvard University Press.

Brazier, F.M.T., C.M. Jonker and J. Treur (2002) Principles of component-based design of intelligent agents. *Data and Knowledge Engineering* 41: 1–27.

Castelfranchi, C. (2001) The theory of social functions: challenges for computational social science and multi-agent learning. *Journal of Cognitive Systems Research* 2: 5–38.

Cederman, L.E. (2005). Computational models of social forms: advancing generative process theory. *American Journal of Sociology* 110(4): 864–93.

Cohen, P.R. and H.J. Levesque (1990) Intention is choice with commitment. *Artificial Intelligence* 42: 213–61.

Cole, M. and S. Scribner (1974). *Culture and Thought: A Psychological Introduction*. New York: John Wiley and Sons.

Collectif Commod (2005) La modélisation comme outil d'accompagnement. *Natures Sciences Sociétés* 13: 165-8.

Conte, R., and C. Castelfranchi (1995) *Cognitive and Social Action*. London: UCL Press.

D'Aquino, P., C. Le Page, F. Bousquet and A. et Bah (2003) Using self-designed role-playing games and a multi-agent system to empower a local decision-making process for land use management: The SelfCormas experiment in Senegal. *Journal of Artificial Societies and Social Simulation* 6(3). Available at http://jasss.soc.surrey.ac.uk/6/3/5.html.

Dray A., P. Perez, N. Jones, C. Le Page, P. D'Aquino, I. White and T. et Auatabu (2006) The AtollGame experience: from Knowledge Engineering to a Computer-assisted Role Playing Game. *Journal of Artificial Societies and*

Social Simulation 9(1). Available at http://jasss.soc.surrey.ac.uk/9/1/6.html

Eco, H. (1997) *Kant and the Platypus*, edn 2000. London: Vintage, Random House.

Ferber, J. (1999) *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. New York: Addison-Wesley.

Funtowicz, S.O. and J.R. Ravetz (1993) Science for a Post-Normal Age. *Futures* 25(7): 739–55.

Gilbert, N., and K.G. Troitzsch (1999) *Simulation for the Social Scientist*. Buckingham, PA: Open University Press.

Granath, J.A. (1991) *Architecture, Technology and Human Factors: Design in a Socio-Technical Context*. Göteborg: Chalmers University of Technology, Architecture Workspace Design.

Hagmann, J. R., E. Chuma, K. Murwira, M. Connolly and P. Ficarelli. (2002) Success factors in integrated natural resource management R&D: lessons from practice. *Conservation Ecology* 5(2): 29. Available at http://www.consecol.org/vol5/iss2/art29

Hare, M. and P. Deadman (2004) Further towards a taxonomy of agent-based simulation models in environmental management. *Mathematics and Computers in Simulation* 64: 25–40.

Hogg, L.M.J. and N.R. Jennings (2001) Socially intelligent reasoning for autonomous agents. *IEEE Trans. on Systems, Man and Cybernetics Part A* 31(5): 381–99.

Jager, W. and M. Janssen (2003) The need for and development of behaviourally realistic agents. In J.M. Sichman, F. Bousquet and P. Davidson (eds), *Multi-Agent Simulation II* (Proceedings of the MABS 2002 Workshop, Bologna, Italy), pp. 36–50. New York: Springer Verlag.

Janssen, M.A. (ed.) (2002) *Complexity and Ecosystem Management: The Theory and Practice of Multi-Agent Systems*. Cheltenham: Edward Elgar Publishers.

Kahnemann, D. and A. Tversky (2000) *Choices, Values and Frames*. Cambridge: Cambridge University Press.

Kaplan, F. (2001) *La naissance d'une langue chez les robots* . Paris: Hermès Sciences Publications.

Kluver, J., C. Stoica and J. Schimdt (2003) Formal Models, Social Theory and Computer Simulations: Some Methodical Reflections. *Journal of Artificial Societies and Social Simulation* 6(2). Available at http://jasss.soc.surrey.ac.uk/6/2/8.html

Lansing, J.S. (2003) Complex adaptive systems. *Annual Review Anthropology* 32: 183–204.

Lissack, M.R. and K. Richardson (2001) When Modeling Social Systems, Models the Modeled: Reacting to Wolfram's 'A New Kind of Science'. *Emergence* 3(4).

Lynam, T., F. Bousquet, C. Le Page, P. d'Aquino, O. Barreteau, F. Chinembiri and B. Mombeshora (2002) Adapting science to adaptive managers: spidergrams, belief models, and multi-agent systems modeling. *Conservation Ecology* 5(2): 24. Available at http://www.consecol.org/vol5/iss2/art24

Maturana, H. and F. Varela (1980) *Autopoieisis and Cognition: the Realization of the Living*. Boston: D. Reidel.

Minsky, M. (1985) *The Society of Mind*. New York: Simon and Schuster.

Oliver, P. (1993) Formal Models of Collective Action. *Annual Review of Sociology* 19: 271–300.

Papert, S. (1980) *Mind-Storms: Children, Computers, and Powerful Ideas*. New York: Basic Books.

Peirce Edition Project (eds) (1998) *Peirce, C.S., The Essential Peirce, Selected Philosophical Writings, Vol. 2 (1893–1913)*. Bloomington and Indianapolis: Indiana University Press.

Tesfatsion, L. (2001) Introduction to the special issue on agent-based computational

Economics. *Journal of Economic Dynamics and Control* 25: 281–93.

Varela, F.J., E. Thompson and E. Rosch (1991) *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.