

Intelligent Social Learning

Rosaria Conte

National Research Council, Institute of Psychology, V.LE Marx 15, 00137 Roma.
Division "AI, Cognitive and Interaction Modelling"

&

University of Siena - Communication Sciences - "Social Psychology"
PSS (Project on Social Simulation) - voice:+39+06+86090210;fax:+39+06+824737
email: rosaria@pssc2.irmkant.rm.cnr.it - http://ip.rm.cnr.it

Abstract

One of the cognitive processes responsible for social propagation is social learning, broadly meant as the process by means of which agents' acquisition of new information is caused or favoured by their being exposed to one another in a common environment. Social learning results from one or other of a number of social phenomena, the most important of which are social facilitation and imitation. In this paper, a general notion of social learning will be defined and the main processes which are responsible for it, namely social facilitation and imitation, will be analysed in terms of the social mental processes they require. A brief analysis of classical definitions of social learning is carried on, showing that a systematic and consistent treatment of this notion is still missing. A general notion of social learning is then introduced and the two main processes which may lead to it, social facilitation and imitation, will be defined as different steps on a continuum of cognitive complexity. Finally, the utility of the present approach is discussed. The analysis presented in this paper draws upon a cognitive model of social action (cf. Conte & Castelfranchi, 1995; Conte, 1999). The agent model which will be referred to throughout the paper is a cognitive model, endowed with mental properties for pursuing goals and intentions, and for knowledge-based action. To be noted, a cognitive agent is not to be necessarily meant as a natural system, although many examples examined in the paper are drawn from the real social life of humans. Cognitive agents may also be artificial systems endowed with the capacity for reasoning, planning, and decision-making about both world and mental states. Finally, some advantages of intelligent social learning in agent systems applications are discussed.

1 Introduction

Two main theses are presented in this paper regarding the phenomenon of imitation:

1. Imitation needs to be modelled in cognitive terms. Far from being a merely behavioural notion, the phenomenon of imitation can be defined in a specific and useful way only if its mental correlates are detected.
2. Imitation is a special case of social learning, which in turn plays a role in behavioural and cultural transmission. In this paper, the social propagation of behaviours and culture is seen as a complex phenomenon which may be realised by means of either social cognitive or non-cognitive processes (see fig. 1).

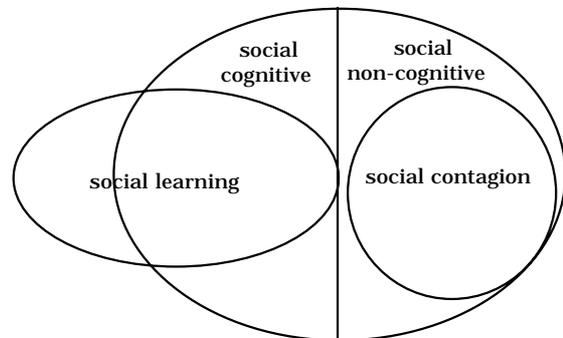


Fig.1: Social Propagation

By a *cognitive* process, it is meant a process (a sequence of operations, such as reasoning, decision-making, etc.) upon symbolic representations (e.g., goals and beliefs). A *social cognitive* process is accomplished upon social mental representations (e.g., social reasoning means reasoning upon social beliefs). Finally, a *social mental representation* mentions another agent and possibly one or more of her mental states (for a discussion of these notions, see Conte & Castelfranchi, 1995, and Conte, 1999).

One¹ of the cognitive processes responsible for social propagation is *social learning*, broadly meant as the process by means of which agents' acquisition of new information is caused or favoured by their being exposed to one another in a common environment.

Social learning results from one or another of a number of social phenomena, the most important of which are social facilitation and imitation (see fig. 2²).

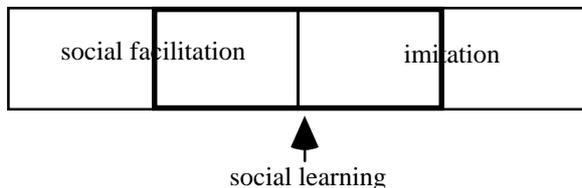


Fig. 2: Social learning

In this paper, a general notion of social learning will be defined and the main processes which are responsible for it, namely social facilitation and imitation will be analysed in terms of the social mental processes they require.

The rest of the paper will be organised as follows. In the next section, where some classical definitions of social learning are analysed, a systematic and consistent treatment of these notions is shown to be missing. In the successive section, a general notion of social learning is introduced and the two main processes which may lead to it, social facilitation and imitation, will be defined as different steps on a continuum of cognitive complexity. In the final section, the utility of a the present approach will be discussed.

The analysis presented in this paper draws upon a cognitive model of social action (cf. Conte & Castelfranchi, 1995; for a synthesis, see Conte, 1999). The agent model which will be referred to throughout the paper is a cognitive model, endowed with mental properties for pursuing goals and intentions, and for knowledge-based action. Therefore, some notions drawn from the formal study of mental states will also be employed.

To be noted, a cognitive agent is not to be necessarily meant as a *natural* system, although many examples examined in the paper are drawn from the real social life of humans. Cognitive agents may also be artificial systems endowed with knowledge and the capacity for reasoning, planning, and decision-making. The interesting question concerning artificial systems is, what are the mechanisms which must be implemented at the agent level to enable them to learn from one

another? Are the mechanisms allowing agents to learn from their physical environment sufficient for them to learn also from or perhaps through their social environment? If not, which additional properties are needed? And, earlier than this, what does social learning mean, which social phenomena are referred to by this notion?

2 Classical definitions

It has been observed (Laland & Odling-Smee, 1999) that the term social learning describes a "ragbag" of heterogeneous phenomena, with a variety of functions. A systematic treatment of these notions is still wanted. In the social psychological literature (Bandura, 1977), social learning is seen as people learning through the observation of attractive and consistent social models. By observing their social models and recording when these apply reinforcing mechanisms, people learn to reinforce themselves (self-reinforcement) to do what others have reinforced, and abstain from doing what others have punished.

This apparently simple and elegant theory has many drawbacks. First, it is exclusively focused on the mechanism of *reinforcement*. What about learning from social models who are unaware about their role and therefore unable to apply prize or penalty?

Second, social learning is essentially meant as a mechanism of *emulation*, which implies the corresponding motivation to look and behave like attractive social models in order to be seen as comparable or similar to them and obtain their approval. What about learning independent of the reputation of the others? Cannot people learn from others without emulating them? Is it possible to formulate a general notion of social learning which includes but is not reduced to emulation?

Finally, given the above notion of social learning, how to distinguish it from imitation? So far, imitation has not been clearly defined. It has been often if not exclusively defined as a behavioural phenomenon. In the typical behaviourist view, recently reworded by Blackmore (1999), imitation is defined as copying a new *form of behaviour*. But what is a new form of behaviour? As a long line of psychological thought has shown (see Plotkin, 1994, for a clear summary), behaviour is essentially a goal-directed or end-directed activity. In this sense, coughing is not behaviour, unless one coughs to signal disappointment or disapproval. When one learns to raise one's arm when meeting another (known) agent, one learns a new behaviour, although the movements involved in such a behaviour were already part of one's action repertoire. In this sense, learning a new *form* of behaviour by imitation means learning a use or meaning (read, goals) which may "in-form" (Plotkin, 1994) a given activity. It then becomes apparent that imitation leads to agents' acquiring novel behavioural information from others, and therefore implies their capacity to draw such information from observed behaviours. Furthermore,

¹ Another important phenomenon is social control, by means of which agents influence one another to comply with intra-groups norms.

² Neither social facilitation nor imitation do necessarily lead to the agent's acquisition of novel information: the former, as often found in the social psychological literature, may lead to improved performance and goal emulation. The latter, especially when aimed to reduce social distance and increase conformity, may lead to comply with standards and norms already acquired by the agent.

some ethologists (see Mitchell, 1987, cit. by Visalberghi & Frigaszy, 1990/99) propose five clauses for a definition of animal imitation:

- a) something C (the copy of the behaviour) is produced by an organism;
- b) where C is similar to something else M (the Model behaviour);
- c) observation of M is necessary for the production of C (above baseline levels of C occurring spontaneously);
- d) *C is designed to be similar to M*;
- e) the behaviour C must be a novel behaviour, not already organised in that precise way in the organism's repertoire.

According to this model, imitation occurs when a given behaviour is designed to be similar or modelled upon that shown by a given target. In other words, a given agent is assumed as a model for several and sometimes convergent reasons (observing the consequences of that behaviour on the target's survival, assuming the target as a social model, obtaining approval from the target, etc.). How to account for such an intuition in an explicit and systematic way? What does it imply in mental terms, which cognitive properties, representations, etc., are required by this notion of imitation?

To sum up, what is needed is

- a) A clearly defined core notion of social learning, which accounts not only for (i) emulation-based processes, but also for (ii) other processes (e.g., purely instrumental social learning); and (iii) is not based only on the reinforcement laws.
- b) A number of specific models corresponding to imitation, social facilitation, etc.; as will be shown, in order to distinguish among these phenomena, the mental states and processes involved must be modelled.
- c) An account of the intuition that, at least in imitation, the behaviour is intended or at least designed to be similar to the model's.

In this paper, an attempt to meet these requirements is made. Rather than a comprehensive theory, what is offered below is a preliminary model of a complex phenomenon, leading from observation of others and their behaviours to acquisition/transmission of these behaviours and the associated beliefs.

3 Behaviour and Mental Representations

One of the most important aspects of social life and intelligence is the diffusion of mental representations (e.g., beliefs) among agents.

Here, it will be said that there is transmission of mental representations through behaviours, when a given behaviour is "interpreted" as based upon some (even very simple) beliefs, goals, etc., which may or may not coincide with those of the observed agent. We will say that mental states leading to the same behaviour are *equifunctional*. A crucial aspect of this view of behavioural transmission lies in the role played by

equifunctional mental states. One could argue that mental states are irrelevant in a model of behavioural transmission, since what is important is their role, not their content: one can assume that if the same behaviour spreads over a given population, the underlying mental states must have been at least equifunctional, if not identical, and that is all one needs to say about them. Below, this view is challenged thanks to a fundamental analytical argument: a theory of mental processes is necessary to understand/predict social phenomena which involve mentally complex systems, such as humans. Let us see why in some detail.

a) A theory of cultural transmission should account not only for *the outcomes* of the process, but also for *the process* itself, and *a model of the process of cultural transmission implies a model of the mental states involved* in it. Democratic institutions (objective outcome) tend to spread at the world-wide level. The problem is how and why this happens, whether under the effect of economic needs (free-market) or thanks (also) to widening *expectations* and *claims* (mental process) for the acknowledgement of human rights.

b) Mental processes which are equifunctional with regard to short-term effects of behavioural transmission may have different long- or mid-term effects. Internet access spread thanks to different mental processes, for example, because it is expected to reduce the costs of exchange, or because it is expected to favour cooperation. Both types of mental processes could have promoted Internet diffusion, but only the latter interpretation will ultimately favour the spread of a "participatory" or communitarian use of Internet (for example, civic networks).

c) In order to predict or account for the stability of specific cultural or behavioural items, one needs to investigate what are the underlying reasons or interpretations. Suppose that, in the daylight, a car proceeding on the opposite side of the street flashes while approaching you. If you interpret this as a greeting, the chances that you will reproduce the same behaviour while approaching other cars are probably not very high. But if you find that an automatic speed limit controller is situated some meters ahead, the chances that you re-interpret the previous driver's behaviour as a convention (e.g., drivers informing one another about speed limit control) increase, and consequently the chances that you adopt the same behaviour with other drivers increase accordingly.

d) A model of the mental processes involved in cultural transmission is fundamental to enable artificial systems to learn from (natural or artificial) others (this is particularly important in infosocieties, which are structurally hybrid multi-agent systems). Which properties should agent systems in a context like e-commerce have to be able to accept useful social laws or conventions (e.g., respect privacy or decency), and avoid or resist useless or undesirable social influence (for example, don't cheat if this results in a loss of reputation of your client's reputation) to control and select external sources of information? Of course, one can implement rules and conventions as simple

machinery, i.e., as action constraints. But in such a case, how to safeguard the autonomy and flexibility of artificial systems? How to enable them to violate social norms and conventions in all the circumstances in which violation is auspicious (for example, in order to comply with more important but incompatible ones)?

4 Social Propagation of Behaviours

Social transmission of behaviours may not imply the transmission of representations, (see the examples in List A of the Appendix to this paper). Think of the spread of the behavioural expression of emotions, which abound in everyday life (Freedman & Perlick, 1979). This actually falls in the wide and generic category of behavioural contagion, which has been explained in terms of two different mechanisms (see Marsden, 1998): social learning, as described in § 2, and "social release" (Ritter & Holmes, 1969; Wheeler, 1966; Levy & Nail, 1993; for a recent analysis, see again Marsden, 1998). Social release essentially consists of a mechanism by means of which, in presence of others, individuals release behaviours which belong to their repertoire but which were inhibited. Both groups of theories, indeed, fail to capture the main difference between contagion and other processes of propagation: the social learning theories do not account for any such difference; the social release theories reduce this difference to a strictly behavioural difference: a behaviour which spreads through contagion is already in one's repertoire, whilst a learned behaviour does not yet belong to one's repertoire. This view has been already criticised in § 2³. Finally, the social contagion is sometimes meant in the rather broad sense of social propagation (Reber, 1995; Marshall, 1994). For example, it is unclear what is meant by "suicidal contagion" (cf., for example, Phillips, 1974). However, the spread of suicide is a rather complex phenomenon which may be due to several mechanisms including but not reduced to contagion.

Behaviour can spread also through social cognitive processes (see List B in the Appendix), mediated by the agents' social goals and beliefs, and by their social competence (e.g., their social reasoning capacity). The essential difference between the Lists A and B lies in the role played by the agents' mental processes in the spread of a given behaviour. Unlike List A, List B contains several interesting examples of social learning. What is needed is a notion of social learning which accounts for the examples in List B, but rules out those appearing in List A.

5 Social Learning

³In addition, the social release theories appear to be based upon a rather unwarranted assumption, namely that any behaviour that is not executed, is inhibited or restrained, and that the social contagion allows for restraint release.

In this section, a subset of phenomena of propagation is addressed, namely those by means of which (a) given (set of) of mental representation(s) (explicit beliefs or rules, procedures, etc.) propagates from one agent to another, and only as a possible consequence, the behaviours that are based upon them are reproduced. This takes place among socially situated agents, which share a common environment and are therefore likely to observe one another.

As announced in the Introduction, social learning is seen here as a process of learning caused or favoured by the agents being situated in a common environment and observe one another. In this sense, the other is not only (perceived as) a criterion for comparison and self-evaluation, but also as a more neutral source of information, which may help or speed several forms of instrumental learning. In addition, social learning is here seen as a multilevel phenomenon, and is defined in a gradualist way starting from an elementary notion. In a minimal sense,

social learning is the phenomenon by means of which a given agent (the learning agent) updates its own knowledge base (adding to, or removing from it a given information, or modifying an existing representation) by perceiving the positive or negative effects of any given event undergone or actively produced by another agent on a state of the world which the learning agent has as a goal.

In the following section, the examples discussed will help clarify this definition.

5.1 Social Facilitation

Here, a general⁴ perspective on social facilitation is taken, allowing to include the notion of local or stimulus enhancement as suggested by the animal ethological literature (for a review, see Heyes & Galef, 1996). For ethologists, local or "... stimulus enhancement refers to a process in which one animal directs another animal's attention to a location (or object) in the environment" (Laland & Odling-Smee, 1999: 5).

More generally, in some species (included humans), one's acquisition of a given piece of information about the environment may be caused by another agent although this does not necessarily imply the propagation of such an information from the mind of the latter agent to the mind of the former. Consider the case in which a given agent, repairing from the rain under a tree, is struck by lightning. An accidental observer will learn something new from her fellow, that

⁴Social psychologists (like Allport and Triplett) refer to a rather restrictive notion of social facilitation, conceived of as the impact of the social context on the quality of one's performance. Thanks to such an impact, a given agent performs better when she is exposed to others performing the same behaviour. Again, according to this definition, social facilitation is based upon emulation and social comparison processes.

is, "never stop under a tree when it rains". In such a case, one agent learns from another without the latter's behaviour to propagate. The example shows that social facilitation is very close to an even more elementary type of learning: the observer might infer the same lethal effect even by watching the tree being struck by lightning. However, as shown by the ethological literature, learning is enhanced by observing the effects of actions or external events on *conspecifics*: what happens to a conspecific will (be expected to) concern me more than what happens to a tree. However, this point deserves further clarification: it is not only the outcome of the process (struck by lightning) that which is bound to elicit the learning process (don't stop under a tree when it rains), but also the *process leading to interpret* the outcome as relevant for oneself (the entity observed).

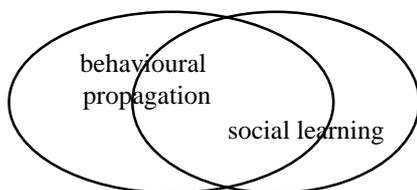


Fig. 3: Social learning with and without behavioural propagation

Social facilitation is a very elementary type of social learning, in which the beneficiary does not necessarily attribute the other any goals or other mental states. In the above example, the same effect might have been probably achieved by the observer, had it seen a piece of wood, rather than a fellow agent, struck by lightning. Of course, if the input comes from an agent, the stimulus to the observer's inference and the chance that she gets new information out of it are higher (since it is higher the probability that the event observed has effects on the observer's goals, which may overlap with the input agent's to some extent).

There are different types of social facilitation phenomena, according to the role played by the input agent (which will be called, the Source) in the Observer's learning process and in its representations. In social facilitation, S may operate as a

a) *pointer* or "*bookmark*": S acts in such a ways as to increase the chances that O perceives a given event, which triggers O's learning process. This is essentially what ethologists call *local enhancement*. As an example, while running after S, O discovers a new region which she⁵ had never realised before. Here, there is no need for S to transmit his own beliefs to O (S might simply escape from O in an unknown direction). O, in her turn, does not acquire a new piece of knowledge by reconstructing S's mental states, nor by

observing the effects of a given event on S's fate. S is a mere accidental cause of O's discovery. S acts as a sort of bookmark or pointer. Here, O learns a new section of the world map.

b) *Qualifier*: S's features may characterise a given environment, and help O characterise or identify it. Suppose I am in a foreign country and badly need a restroom, but cannot tell from the written signs which is the ladies' and which is the gentlemen's toilet. One possible solution is wait and see which way will take the next newcomer of either gender. Interestingly, the social cognitive process which occurs is the same but leads to alternative behaviours: if the newcomer belongs to my gender I will act alike; if he is of the opposite gender, I will take the alternative way.

c) *Activator*. This is shown by the example of milk bottle top opening in British tits (Hinde & Fisher, 1951). Let us see how Laland and Odling-Smee (1999: 6) discuss this example: "These birds learned to peck open the foil cap on milk bottles... . Hinde and Fisher found that this behaviour probably spreads by local enhancement, where the tits' attention is drawn to the milk bottles by a conspecific, and after this initial tip off, they subsequently learn on their own how to open the tops". However, Hinde and Fisher's explanation is insufficient. The learning process is facilitated by S in a double way: S draws O's attention on a given object, which possibly "activates" O's goal of manipulating it, and therefore leads O to exhibit the same behaviour as S. Here, propagation occurs. However, there is no need that O actually represents S as a "manipulator", nor, a fortiori, that O attributes S any capacity or mental states. S points to a new object which might activate a built-in routine for manipulation. An analogous example is offered by the acquisition of dietary preferences among rats (Galef, 1996; see again, the discussion of this example in Laland & Odling-Smee's, 1999: 6), which prefer "to eat foods that other rats have eaten".

d) *Belief-holder*: a subset of S's inferable beliefs may help O to identify and understand the environment. An interesting example of this phenomenon is offered by people's recognising a given (social) setting by observing others' behaviour: if someone is standing on the edge of the sidewalk, it is probably there where the bus stops. In such cases, O resorts to her pre-established beliefs about S (pedestrian): people standing up motionless in the street usually are waiting for someone or something. Interestingly, O may have a pre-existing goal (taking the bus), which S helps her to achieve by marking how to verify its preconditions (find the place where the bus stops). Alternatively, this goal may be activated by O's perception of S's behaviour and by inferring the associated mental states (O is walking to destination, but since she understands that a bus-stop is near, she may get on the next bus). In such a case, social facilitation allows for social propagation: a given (set of) belief(s) travels from S to O. Indeed, O decodes S's beliefs from his behaviour and incorporates them into her knowledge base (unless she finds evidence that S is wrong or her inference is incorrect).

⁵In the remaining of the paper, explicit reference will be made to human agents to facilitate the reader's understand the reference of pronouns (O will be a female agent, and S will be a male agent). However, some, if not all the processes analysed may occur among non-human organisms and even among non-natural systems.

e) *Experimental "testbed"*: this is shown by the example of looking for a shelter from rain. By observing what happens to S, O learns to avoid trees. Here, O learns a negative effect of a known plan of action. Examples of this sort abound in social life: agents not only observe and learn given behaviours from one another, but also avoid the costs of a direct experiment, and learn the positive or negative (side-)effects of current plans/procedures etc.

f) *Subject of norms, standards, conventions*: S's behaviour may indicate existing standards, norms, and conventions. Independent of whether O will decide to accept or reject them, and of whether to comply with them or not, others may be a fundamental source of information about formal or informal norms, customs, habits and any other factor of regulation of one's (social) conduct. More basically, think of O as an external observer, an anthropologist or ethnographer. She may learn a lot about how a given society is organised, differentiated, what are its social hierarchies, etc. from the behaviours of the society's members. But even independent of standards and norms, O may learn a lot about social categories, reputation, roles, etc. by observing how agents interact with and react to one another.

To sum up, social facilitation is a mechanism by means of which a given agent updates her knowledge base, including social and pragmatic knowledge, by observing others, their features and behaviours, and possibly (but not necessarily) by inferring their mental states.

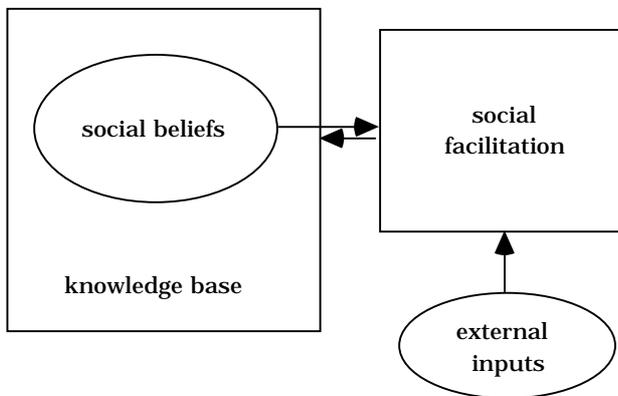


Fig. 4: Inputs and outputs of social facilitation

5.2 Imitation

In the previous section, social facilitation has been defined and described as a type of social learning in which the learning agent (O) updates her knowledge base by perceiving the relationship between another agent (S) and its physical or social environment. Such relationships may (or may not) include the effects of S's behaviour on the environment, and/or the effects of the environment on him (and possibly his achievements). In social facilitation, O receives information relevant for

her current or potential goals by observing S in a common environment. Consequently, O forms beliefs/perceptions about S (from which she acquires novel information), but her goals do not mention S. S simply plays the role of an implicit, undeliberate, even accidental indicator or even informant about the environment.

In this section, another step of the social learning process is analysed, namely imitation. Imitation is here defined as a phenomenon of social learning in which the learning agent is ruled by two social *goals* concerning S (a social goal being defined as a goal which mentions another agent's mental states; cf. Conte, 1999):

a) know what S does, how he behaves, how he looks, etc. in order to find out standards, rules, or simply means to achieve her own goals. O's social goal is a means for O to reach another goal of hers. The latter might be specific or generic. For example, O may not know how to use the silverware in a fancy restaurant. She then looks around to see what her fellows do with them.

b) adopt S's goals and/or other mental states and possibly the consequent behaviours, *as long as* O believes that S is an appropriate or adaptive model in a given domain. In the formal treatment of mental states, a goal *relative* (Cohen & Levesque, 1990) to a given belief is a persistent but conditioned goal, that is, a goal which is pursued as long as it is found either unfeasible or already achieved (persistent) or unless the belief associated to it is revised or retreated (conditioned). In the case of imitation, the goal is relative to O's *social* belief: O imitates another agent as long as she believes that it is useful and convenient to do so, namely as long as the other shows an appropriate or adaptive behaviour, looks, style under given circumstances.

Imitation is a behaviour ruled by the goal that a given agent (O) be-like or act -like another agent M (which stands for Model), as long as M is (perceived as) a suitable model under a given circumstance.

The main difference between social facilitation and imitation is that in the former case, O has social beliefs about S, from which O obtains relevant novel information. In imitation, instead, O pursues a number of social goals with regard to M, relative to her belief that M is a good model. These goals actually suggest interesting operational criteria for a model of imitation: if a system is ruled by a goal *relative to* a given belief (Cohen & Levesque, 1990), the system will have to (a) check the truth value of its current belief - in the case of imitation, it will repeatedly monitor (i) M and his doings, (ii) how good (e.g., adaptive or successful) M is as a model; (b) the relativized goal is a persistent but conditioned one - in our case, O will persist in imitation as long as she believes M *is* a good model.

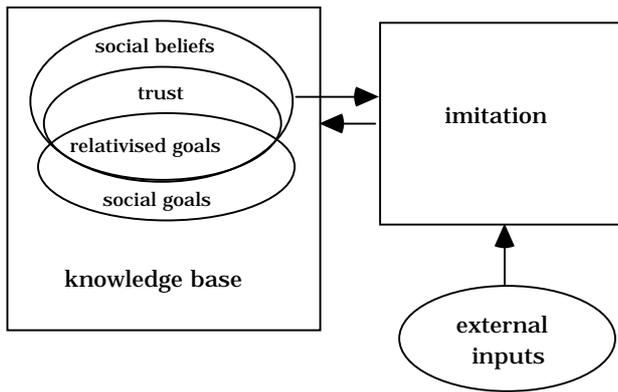


Fig. 5: Mental States in Imitation

5.2.1 Imitation: Goal-Directed or Goal-Oriented Behaviour?

The preceding definition might raise an important question: what type of goals are implied by imitation? What does it mean that O is ruled by a set of social goals? To what extent can this notion of imitation be used, which systems can it be referred to? More explicitly, which mental capacity (or complexity) is required for a system to exhibit and be attributed this type of behaviour? What is the relationship between Mitchell's clause (4) (that behaviour C is *designed to be similar to M*) and the definition here provided?

The question is a difficult one, which will find no conclusive answer here. However, it is at least necessary to recall that goal-directed behaviour (Plotkin, 1994, in line with the Piagetian definition) includes goal-governed behaviour (cf., Conte & Castelfranchi, 1995, ch. 8), that is, behaviour selected by the evolutionary process *to achieve* a given effect, without this effect being necessarily represented as an explicit goal in the imitator's mind. Consequently, imitation among some species might have been selected as one agent's *monitoring of given others* (e.g., parents) in order to *be-like or act-like them* as long as they are (perceived as) *good models* under given circumstances. What is a good model in evolutionary terms? How can "good" models be biologically transmitted without bringing into play a sort of Lamarckian evolution? This is a difficult question which ethological literature can help to answer. For many species, any conspecific has been selected as a good model (read, better than a non-conspecific) at least relative to given types of behaviour (e.g., dietary rules). For other species and in other behavioural contexts, adults have been selected as "good" models (infant chimps imitating adults foraging for termites using stalks), or more specifically "first-perceived" adults (imprinting). To adopt the dietary preferences of conspecifics, in such analysis, is attributable to imitation if a set of operational criteria, corresponding to the relativized goals mentioned above, are met. To define such criteria is beyond the scope of the present model. However, the notion of relativized goal can contribute to formulate them. A behaviour governed by a relativized goal is a behaviour persistent but conditioned to a given belief (that M is a good

model): (a) imitating different sets of models (e.g., conspecifics Vs parents) for different domains (dietary preferences vs. problem-solving heuristics) is a necessary but insufficient criterion; in order to check conditioned persistence, one should probably (b) check whether the animal persistently monitors its model (like for example, in imprinting), (c) check whether imitation is conditioned to the model's success (e.g., abandon a given diet learned from a conspecific after the latter's death).

5.2.2 Why Imitate...

There are several main reasons for O to imitate M,

a) Know what to do: in many (social) circumstances, agents may ignore the efficient or more convenient procedures, rules, plans to achieve their goals. Sometimes, they may even ignore which goals can be achieved, or which activities are feasible or safe, etc.. For example, while taking a walk in an unknown city, I follow the main stream of pedestrians simply because I do not know where to go to, which neighbourhood is safe enough, etc..

b) Comply with social standards and norms. This is very close to social facilitation in the sense that O observes others to infer information about the world from their behaviours. But what is characteristic about imitation is that O is interested not only to find out what the norms are, but also to see how others react to them, to what extent they keep them into account, which ones are applied and which ones are instead ignored. By this means, O learns both relevant *information* about the (social) world, and relevant (social) *conducts*.

c) Fulfil given roles. Sometimes, imitation is prescribed. For example, parents are models that children *must* follow (this has obvious biological forerunners, such as imprinting).

d) Compensate one's inadequacy, ignorance, inexperience in problem solving. If I do not know how to use the chopsticks, I will probably take the customers of a Chinese restaurant, or my friend Yang, as suitable models to observe carefully in order to reproduce their behaviours.

e) Avoid complex problem solving or risks. Sometimes, I am too uncertain about the consequences of a given conduct, and rather than sustaining the cost of a direct experience, I decide to shelter under someone else. For example, if I am alone, I may stop on the edge of a slope, but I may risk to follow someone who decides to proceed.

f) Reduce social distance. At times, people imitate others to conform to them, thereby reducing the social differences within the group. In these cases, O's goal to adopt M's goals, preferences and behaviours is not relativized to her belief that M is an appropriate model, but rather to her belief that M persists in having those goals, preferences, etc.. In its most extreme form, this leads to O following M whatever he does.

g) Share responsibility, commitment, etc. and their costs. In social impact theory, a crucial factor of explanation of bystanders' intervention in emergencies

is precisely the existence of other bystanders. Indeed, the more they are, and the less cohesive the group, the less efficient or timely the intervention of bystanders during emergencies. Why? As Latané and Darley (1970) convincingly explained, people want to share the responsibility of both the decisions involved (whether or not to define the episode as an emergency, and whether or not to intervene) with others. No one wants to find herself in a "vulnerable", isolated, position. Consequently, each one checks what others do and how they react. Of course, this generates a bottleneck: each waits for another to make the first move, which no one will then undertake.

Generally speaking, imitation appears as a short-cut in problem solving and planning. O minimises the costs she should otherwise invest in these activities by accepting others' outputs. Of course, a trade-off may be envisaged here: on one hand, O reduces her own costs, on the other she might increase them by following others' conducts which later might appear useless or risky. Indeed, imitation implies delegation and ultimately trust: O implicitly delegates others to do (part of) the job she should do. She must trust M to some extent. Consequently, imitation leads to another, intrinsically social type of problem-solving and reasoning: whom to trust, and about what? How to tell when someone is trustworthy or reliable, how to tell that his conduct is adaptive and that it is then reasonable to follow it? To put it otherwise, when do agents perceive themselves as adequate in problem solving and when, instead, do they prefer to delegate this to others?

5.2.3 Whom to Imitate

This question is closely related to the role of trust: O imitates M when she trusts M. But to what extent should she trust him, with regard to which competencies or characteristics?

First, imitation may have more or less domain-specific target. This is because trust is relative to specific contexts and domains of competence: I will certainly look at Yang in using chopsticks, but may have no good opinion of his command of the English language and therefore refrain from imitating him in such a context. In addition, imitation may be individualised or not: I may decide to look at my friend Jenny rather than John, because all considered, I trust her competence, problem solving capacities, etc., to a higher extent than I trust his. On the other hand, I may want to look at any colleague who obtained a promotion in the last two years.

Another important dimension, which is related to trust but different from it, is the goal of imitation: a youngster will feel more likely to find models in her own age co-hort than in others. Here, the goal is not problem-solving but reducing one's social distance from a given social aggregate. Therefore, the target of imitation is any agent who is a good representative, a typical exemplar of that aggregate. Obviously, prototypes are *trusted* to possess the characteristics which are essential to the category of reference.

Finally, imitation may be based upon observable frequencies: in many cases, the more frequent a given behaviour and the more it is target of imitation. This has at least three reasons:

a) first, the more frequent a given behaviour, the more it is perceived as rational, in the sense of independent of subjective, idiosyncratic preferences and biases

b) the more frequent a given behaviour and the more it is perceived as one which as proved the fittest (selected by success)

c) the more frequent a given behaviour and the more it is perceived as prescribed, or even mandatory.

5.2.3 ... and What to Imitate

Unlike the classical view of imitation as a strictly behavioural notion, imitation is here seen as a special case of intelligent social behaviour, in which the Observer intends or is designed (to use Mitchell's phrase) to be similar to a given Model, by adopting M's

a) Behaviours; here, it is important to recall that imitation does not necessarily mean to learn a new set of "movements", but rather learn to give a new meaning (and also a new context) to a given behaviour. As Laland and Odling-Smee (1999: 6) observe, by referring to Heyes (1995), "... it is *not the motor pattern* that is learned, but rather existing *topographically defined behavioural elements*, alone or in combination, are associated with the consequences of the behaviour, in a particular context" (italics are mine). More explicitly, in imitation, agents learn to adapt their behaviour to achieve new goals.

b) Internal states, including the mental ones, such as beliefs, values, preferences (think of the dietary preferences among rats), goals, practical heuristics (think of the washing of sweet potatoes among Japanese macaques, Heyes & Galef, 1996). Internal states should not be confused with internal behaviours, that is, mental actions and operations, although these may also be targets of imitation.

c) Skills (think of Goodall's, 1964, well-known example of the skills necessary for foraging for termites using stalks acquired by infant chimps imitating adults).

d) External standards and criteria, which are inferred to (i) input A's behaviour, (ii) be mirrored in A's behaviour, and (iii) rule it.

The things that are imitated are (either learned cognitively or selected via biological evolution) relevant to agents' adaptation. Such relevance assumptions are essential if imitation (1) is to be at all possible and (2) will combine efficiency with effectiveness.

5.2.4 How to imitate

The mental process required by imitation is variable and ranges from the blind reaction of a baby duck following the first mobile object occurring in its perceptive field (which may happen to be an ethologist rather than its

mother) to a much more complex set of mental operations and representations.

In the case of imprinting, the mental properties required by imitation are rather poor, since the difficulties have been somehow managed at the evolutionary rather than at the individual level. In other words, the mechanism is not based upon by the single agent's mental representations nor allowed by its reasoning capacity. Rather, during the evolution of the species the evolutionary process has gradually selected a sensory-motor schema which allows the individual animal to answer adaptively some questions crucial for its own survival.

In most examples of imitation, and quite often among human adults, imitation requires a rather complex set of mental representations and processes:

a) Social beliefs, i.e. (i) information about M, his social status, mental states, etc.; (ii) information about M's credibility, reliability, expertise, etc.. Imitation implies trusting M (or a set of agents, possibly coinciding with the whole social environment) as a source of information about adaptive behaviour (for an analysis of trust, see Castelfranchi & Falcone, 1998). However, the extent to which trust promotes imitation is variable.

b) Social reasoning, that is the capacity to infer M's goals, beliefs, values, etc. from his behaviours or appearance.

c) Relativized social goals, both the goal to acquire information about M and the goal to be similar to him, as long as he is believed to be a suitable model.

6 Advantages of the present analysis

Since imitation may be displayed even on the ground of built-in schemata and reactive behaviour, what is the use of a cognitive model as one presented in this paper? This question is even more crucial if one does not aim only at describing imitation among natural organisms, but also at implementing imitation in artificial systems: if there is a way to obtain the same result with low-complexity mechanisms (such as routines and production rules), why then bother with high-complexity, cognitive mechanisms?

There are several answers to this question, both at a scientific level and at the level of agent and multiagent systems applications.

6.1 To improve scientific understanding of social learning

It is yet unclear what can be learned via simpler mechanisms, to what extent social learning can be effectively achieved thanks to simpler mechanisms at the level of the agent. Certainly, a model of imitation which does not account for its cognitive ingredients will hardly enable us to distinguish social learning from pure social contagion. The main difference between these phenomena seems to reside precisely in the role played by the agents' mental processes in each of them: in

social contagion, a given behaviour spreads automatically and easily, and often as quickly it decays. In social learning, modifications of the agent's states or behaviours is more robust and durable. The question is how such a difference can be explained and somehow reproduced.

Third, cognitive ingredients allow us to give more adequate and complete accounts of different forms of social learning, e.g., social facilitation and imitation. Indeed, a low-level definition of imitation as a mere behavioural phenomenon does not do justice to the ethological evidence that only animals like apes and dolphins do exhibit imitation, while many others exhibit only simpler types of social learning, such as social facilitation, if any at all. Why should this be the case if imitation were essentially based upon mechanisms such as matching between kinesthetic and visual images, enough elementary, or simple, to be executed by members of lower-level species?

Fourth, a cognitive model allows for an evolutionary, or at least a stepwise view of social learning and intelligence. It allows for different degrees and types of social influence to be investigated and some forerunners of social reasoning (reasoning upon others' minds) to be identified. For example, the capacity to use others as environmental bookmarks requires, and therefore gradually evolves into, the capacity to map the environment by deconstructing how others behave in it. As a consequence, certain forms of social facilitation may require as complex mental processes as those involved by imitation. But it is also the case that imitation represents an evolution of the processes involved in more elementary forms of social learning.

6.2 Socially intelligent agents for technological applications

Technological applications in the field of agents, require more sophisticated models of interactive and social competencies (for an argumentation of this claim, see Conte 1999b). In particular, the necessity to improve agents' capacity to learn from one another is largely shared by agent systems scientists. Attempts at implementing this capacity often draw upon classifier systems, adaptable agents, etc.. Two orders of questions arise here:

(a) How far can one go with the behavioural model of learning allowed by current solutions, such as classifier systems; on the other hand, what are the advantages for agent systems' applications of implementing intelligent social learning?

(b) More crucially, which properties are needed at the level of the agent to implement intelligent social learning?

6.2.1 Why implement intelligent social learning

Current learning systems are essentially stimulus-response systems, either symbolic (e.g. Learning

Classifier Systems, cf. Watkins, 1989) or sub-symbolic (e.g. neural nets).

Evolutionary Reinforcement Learning, Classifier Systems used for adaptive agents (Holland, 1992), allow the acquisition of new (social) beliefs, and the emergence of new strategies and agents (cf. Holland, 1995). Whilst these systems actually implement learning and evolving mechanisms, and have allowed to study the emergence and spread of interesting social phenomena, they do not yet allow to implement:

(a) The acquisition of attitudes, preferences, and other non-behavioural features, which implies that these be implemented at the level of the model, and, moreover, that they are recognised and interpreted by the learning agents.

(b) Selective learning and resistance to change: how to implement at the level of the agent, given criteria for learning (learning what is desirable, fair, respectable, etc.)? This is essential to preserve some degree of system's robustness, and provide the agents with a relative capacity and criteria for resisting external, namely social influence. On the other hand, it allows to implement selective learning, and "desirable global effects" to emerge and spread. Selective social learning is essential to implement the spread of social norms and conventions in multiagent systems.

(c) Social models: these represent an interesting criterion of selective learning, and therefore an enforcement mechanism of conventions and social norms, in which given others (the so-called significant others) are assumed as good, convenient, reasonable, respectable, etc. models for imitation. As an additional advantage, to implement social models would promote the agent-based simulation study of the emergence of social hierarchies and structures (such as coalitions, alliances, etc.).

(d) Different attitudes towards learning: natural agents vary as to their capacity for and attitude to learning. To implement learning variety is essential to several domains of agent systems applications (believable agents, synthetic actors, multiagent systems, etc.), and requires a model of the processes and mechanisms which lead agents to *want* to learn.

6.2.2 How to implement intelligent social learning

To fulfil the tasks listed above, agents need to

(a) acquire social mental representations, that is social beliefs and social goals and intentions, including the goal to imitate others, as well as the capacity to

(b) attribute external and internal features to others, and update or instantiate models of others

(c) reason upon social beliefs, thereby generating new beliefs and take them into account while acting and imitating,

(d) form relativized social goals, that is social goals relative to social beliefs

(e) compare one's own knowledge base with that of others

(g) decide whether to imitate, solving potential conflicts goals among the goal to imitate and not to imitate, according to some criterion

(f) adopt external criteria for selective imitation (e.g., social desirability)

(g) decide which agents to imitate, instantiating social models to existing exemplars.

7 Summary

In section 2, some requirements of an adequate treatment of social learning were identified and found still wanting in the current models. The analysis presented in this paper seems to contribute to meet those requirements.

Both a core notion of social learning and some specific notions relative to the main processes leading to it - social facilitation and imitation - have been provided, which allowed both the similarities and the differences between these processes to be emphasised. This analysis presents two main characteristics:

a) Rather than focusing exclusively on emulation-based processes, and the improvement of one's performance, the *more general* phenomenon of one's acquisition of new information has been addressed.

b) Rather than grounding social learning on social reinforcement, social cognitive properties and mechanisms have been investigated, which seem to account for both the *similarities* and the *specificities* of the two phenomena of interest.

Finally, the utility of the approach presented here has been examined at both the scientific level and at the level of agent system applications.

References

- A. Bandura, *Social Learning Theory*. General Learning Press, New York, 1971.
- S. Blackmore, *The Meme Machine*. OUP, Oxford, 1999.
- C. Castelfranchi, R. Falcone, Principles of Trust for Multi-Agent Systems: Cognitive Anatomy, Social Importance and Quantification, *Proceedings of the International Conference on Multi-Agent Systems (ICMAS 98)*, Paris, La Villette, July 1998.
- P.R. Cohen, H.J. Levesque, *Persistence, Intention, and Commitment*, In P.R. Cohen, J. Morgan, M.A. Pollack (eds), *Intentions in Communication*, 33-71. Cambridge, MA: MIT Press, 1990.
- R. Conte, Social Intelligence Among Autonomous Agents, *Computational and Mathematical Organization Theory*, forthcoming, 1999.
- R. Conte, C. Castelfranchi, *Cognitive and Social Action*. UCL Press, London, 1995.
- J.L. Freedman, D. Perlick, Crowding, contagion and laughter. *Journal of Experimental Psychology*, 15:295-303, 1979.

- B.G. Jr. Galef, Social enhancement of food preferences in Norway rats: A brief review. In *Social Learning in Animals: the Roots of Culture*. C.M. Heyes, B.G. Jr Galef (eds), Academic Press, New York, 49-64, 1996.
- J. Goodall, Tool using and aimed throwing in a community of free living chimpanzees. *Nature*, 201, 1264-1266, 1964.
- C.M. Heyes, B.G. Jr. Galef (eds.) *Social Learning in Animals: the Roots of Culture*. Academic Press, New York, 1996.
- C.M. Heyes, Imitation and flattery: A reply to Byrne & Tomasello. *Animal Behaviour* 50, 1421-24, 1995.
- R.A. Hinde, J. Fisher, Further observations on the opening of milk bottles by birds. *British Birds*, 44, 393-396, 1951.
- M.L. Hoffman, Altruistic behaviour an the parent-child relationship, *Journal of Personality and Social Psychology*, 31, 937-943, 1975.
- K.N. Laland, J. Odling-Smee, The Evolution of the Meme. Paper presented at the Conference "Can Memes Account for Culture?", Cambridge, UK, 4-5 June, 1999.
- B. Latané, J.M. Darley, *The unresponsive bystander: Why doesn't he help?* Appleton, New York, 1970.
- D.A. Levy, P.R. Nail, Contagion: A theoretical and empirical review and reconceptualization. *Genetic, Social and General Psychology Monographs*, 119:235-183, 1993.
- P. Marsden, . Memetics and Social Contagion: Two Sides of the Same Coin? *Journal of Memetics - Evolutionary Models of Information Transmission*, 2, 1998.
- http://www.cpm.mmu.ac.uk/jom-emit/1998/vol2/marsden_p.html
- G. Marshall, (ed.) *Concise Oxford Dictionary of Sociology*. OUP, Oxford, 1994.
- R.W. Mitchell, A comparative-developmental approach to understanding imitation. In P.P.G. Bateson, P.H. Klopfer (eds.), *Perspectives in ethology, vol. 7: Alternatives*, 183-215, Plenum Press, New York, 1987.
- D.P. Phillips, The influence of suggestion on suicide: Substantive and theoretical implications of the Werther effect. *American Sociological Review*, 39:340-354, 1974.
- D.P. Phillips, The impact of fictional television stories on U.S. adult fatalities: New evidence on the effect of the mass media on violence. *American Journal of Sociology*, 87:1340-1359, 1982.
- D.P. Phillips, The impact of mass media violence on U.S. homicides. *American Sociological Review*, 48:560-568, 1983.
- H. Plotkin, *Darwin Machines and the Nature of Knowledge*. Penguin, London, 1994.
- A.S. Reber (ed.), *The Penguin Dictionary of Psychology* (2nd ed.). Penguin, London, 1995.
- E.H. Ritter, D.S. Holmes, Behavioral contagion: Its occurrence as a function of differential restraint reduction. *Journal of Experimental Research in Personality*, 3:242-246, 1969.
- J. Searle, *The Construction of Social Reality*, Penguin, London, 1995.
- R. Tuomela, M. Bonniver-Tuomela, From Social Imitation to Teamwork. In G. Holmström-Hintikka, R. Tuomela (eds) *Contemporary Action Theory*. Vol. II. Kluwer, Amsterdam, 1-47, 1997 .
- E. Visalberghi, D. Frigaszy, "Do monkeys ape?" Ten years after. *Paper presented at the Workshop on "Imitation in Animals and Artifacts"*, 1999.
- L. Wheeler, Towards a theory of behavioural contagion. *Psychological Review*, 73:179-192, 1966.

Appendix

List A

"Black-out" effect, or restriction of the space of possible actions. Here, no social competence operates, but a high regularity, or convergence, in agents' (social) behaviour due to some central extraordinary event. No mutual influence is exercised by the agents undergoing this effect. Still, they converge on the same behaviour (as happens in explosion of the birth rate nine months after a real black-out) thanks to a severe restriction of feasible actions.

Direct exposition, or the "party-shower" effect⁶. After the 1997/98 repeated earth moves in Central Italy, people were reported to develop compulsory paranoid thoughts. The same can be expected to be reported by the Turkish or Taiwan population after the more recent earthquakes in those areas. As in the black-out effect, a major discontinuity had been introduced in their normal life by a non-ordinary event. But unlike the previous effect, in this case, the influence of this event on agents is determined by their perception and interpretation of the event, and by the consequent feeling of powerlessness. However, neither influence nor imitation are (necessarily) at stake: agents did not need to communicate to, nor observe, one another (although, in fact, they most certainly did) for their feelings and behaviours to spread over the whole group.

Behavioural "domino" effect. With this type of effect, we enter in a more interesting sub-area of phenomena, namely transmission (and possibly convergence) due to the non-mental effect of agents' behaviour on, and through, one another. Consider the case in which, in social or public settings (for example, a crowded restaurant⁷), you are obliged to raise your voice otherwise your friends won't be able to hear you. Here, agents do not form any representation of the others nor of their behaviour. They simply raise their voice in order to be audible, thereby causing a corresponding continuous increase of noise⁸.

List B

The social models' influence. The propagation of mental anorexia among young women in Western societies is often considered as a consequence of their exposition to the unhealthy aesthetic standard of the "slender type". Of course, this does not account for the intrinsic replication success of the aesthetic standard in question (which is a memetic effect), but accounts for the width of the phenomenon: young women are strongly and widely influenced by it because fashion models and top girls *are* skinny. (This belongs to the same category of phenomena observed by Phillips, 1982, 1983 in his studies on the impact of media on social violence).

Socially-based goal-activation. Consider Weber's famous example discussed by several authors (for one example, Tuomela & Bonniver-Tuomela, 1997): while walking in the street you realize that people around you have opened their umbrellas. You then almost certainly infer that it is raining, although your thick hair or wide hat prevented you from perceiving the first drops. This inference will activate a goal of yours, i.e. not to get wet. Once such a goal has been activated, the role of the input agents stops. You are able to find a solution on your own: if you have an umbrella (which is already stored in your knowledge base as a good means to avoid getting wet), you will probably follow the example of your neighbours. But if you were not so mindful as to get one, you may decide to hasten your pace, or stop at the next pastry shop, or finally change your mind and get back on your steps. In all these cases, your decisions are influenced by your interpretation of the perceived passengers, but only in the former you actually replicate their behaviours (opening the umbrella).

Elite-oriented conformity. In this case, agents are ruled by their goal to show the same taste and preferences as those shown by (significant) others. They will exhibit given tastes and standards as long as they believe that these are shared by their models. Interestingly, this is complementary to the Simmel effect, shown by agents who consider themselves as "élites": these have the goal of maintaining preferences as long as these are shared only by their affiliates. As soon as others will converge on the same preferences, in order to be perceived as affiliates to the élite, the elitarian agents will drop them and turn to other, more selective, ones; and the process will be re-initialised.

⁶This name is after Searle's (1995) example of the prompt flight of participants at an out-doors party at the first evidence of an incipient shower.

⁷This example was shown to me by my colleague Cristiano Castelfranchi.

⁸This is also known as the "arena" effect: if during the performance, people in the first rows stand up, those who are right behind are automatically induced to follow their behaviour, and so on and so forth until people occupying the farthest seats.

The "vulnerable position" effect. On the highway, if everybody exceeds the speed limit, you are obliged to break the rule in order not to be hit sooner or later from behind. Your behaviour is influenced by the frequential norm established by others. However, neither imitation nor any representation of the other is (necessarily) involved. This is a mere case of an emergent regularity (which results in violating a specific norm).

Automatic contagion of emotion expression. At a party, if one starts to yawn, s/he will most certainly be followed by many participants. If you happen to listen to some foreigners speaking in an incomprehensible language and to see them bursting into laughter, you can't help laughing as well. If asked why you were laughing, you won't be able to give any good reason; still the automatic impact of laughter is irresistible.

The group effect (or social impact). The famous Social Impact Theory (Latané & Darley, 1970) accounts for an interesting variant of the vulnerable position effect in groups of agents facing an emergency. To avoid an isolated and therefore "vulnerable position", each bystander waits for someone else to make the first move and provide help to the victim. As a consequence, no-one will provide the help required.

Emotional sharing. Consider the case of empathy (cf. Hoffman, 1975). In this phenomenon, emotion spreads thanks to a specific mental process. A beggar shows helplessness and even despair because he is helpless (he believes something like "How dreadful: I am helpless"). The empathic passenger will feel sad if she believes "How dreadful: he is helpless". However, thanks to the empathic mechanism (rather mysterious, indeed, in absence of some biological source of solidarity), the passenger shares (to some extent and for a short time) the emotion or feeling expressed by the beggar. Here, something new occurs: the passenger perceives the emotional state of the beggar and infers his/her more general (social) state: empathy is in fact based upon specified attributions. In fact, people do not share the feelings of those who are perceived as responsible for their mishaps. Only under given attributions, they come to share the feelings of the victim. The emotional sharing is therefore caused by an inferential process, by a reasoning applied to the mental and objective conditions of the victim. However, no imitation occurs yet.