THE INVISIBLE (LEFT) HAND:

For a (Pessimistic) Theory of Spontaneous Social Order

A critical homage to F. von Hayek

Cristiano Castelfranchi

***Institute for Cognitive Sciences and Technologies* -Roma**

Agent-based Social Simulation will be crucial for the solution of one of the most hard **theoretical** problems of economic and social sciences:

the spontaneous organization of a "dynamic social order" that cannot be planned, but emerges out of intentional planning agents guided by their own choices.

This is the problem that **Hayek** assumes to be the real reason for the existence of the Social Sciences.

I will examine in particular:

- the crucial relationships between the intentional nature of the agents' actions and their explicit goals and preferences, and the possibly **unintended 'finality' or 'function' of their behavior**.

- in favor of **'cognitive architectures' in computer simulations**.

- propose some solutions about the theoretical and functional **relationships between agents' intentions and non-intentional 'purposes'** of their actions.

- **'Social order' is not necessarily a real 'order' or something *good*** and desirable for the involved agents; nor necessarily the best possible solution.

- **It can be bad** for the social actors against their intentions and welfare although emerging from their choices and being stable and self-maintaining.

- **Hayek's** theory of spontaneous social 'order' and **Elster's** opposition between intentional explanation and functional one will be criticized.

Agent-Based Social Simulation
for the Social Sciences

a new paradigm for SS

systems of equation ?

"swarm intelligence" ?

we need a more complex model of the agent, and precisely *a more complex theory of action* and *models of mind*. (ex. Lomi, Larsen 1995 ; 1996)

It is a wrong move and an illusion that of separating and opposing >> **"emergent intelligence"** (or emergent cooperation) >> **"mental intelligence"** (and deliberate cooperation)

Agent-Based Social Simulation for the Social Sciences

In MAS we risk to have this opposition:

- **reactive agents with collective unconscious problem solving (*emergent functionalities*),**

- **cognitive agents that should base all their cooperation on mutual knowledge, joint intentions, negotiation, awareness and deliberation of their cooperative mechanisms. (**Hyper-Cognitive view**)**

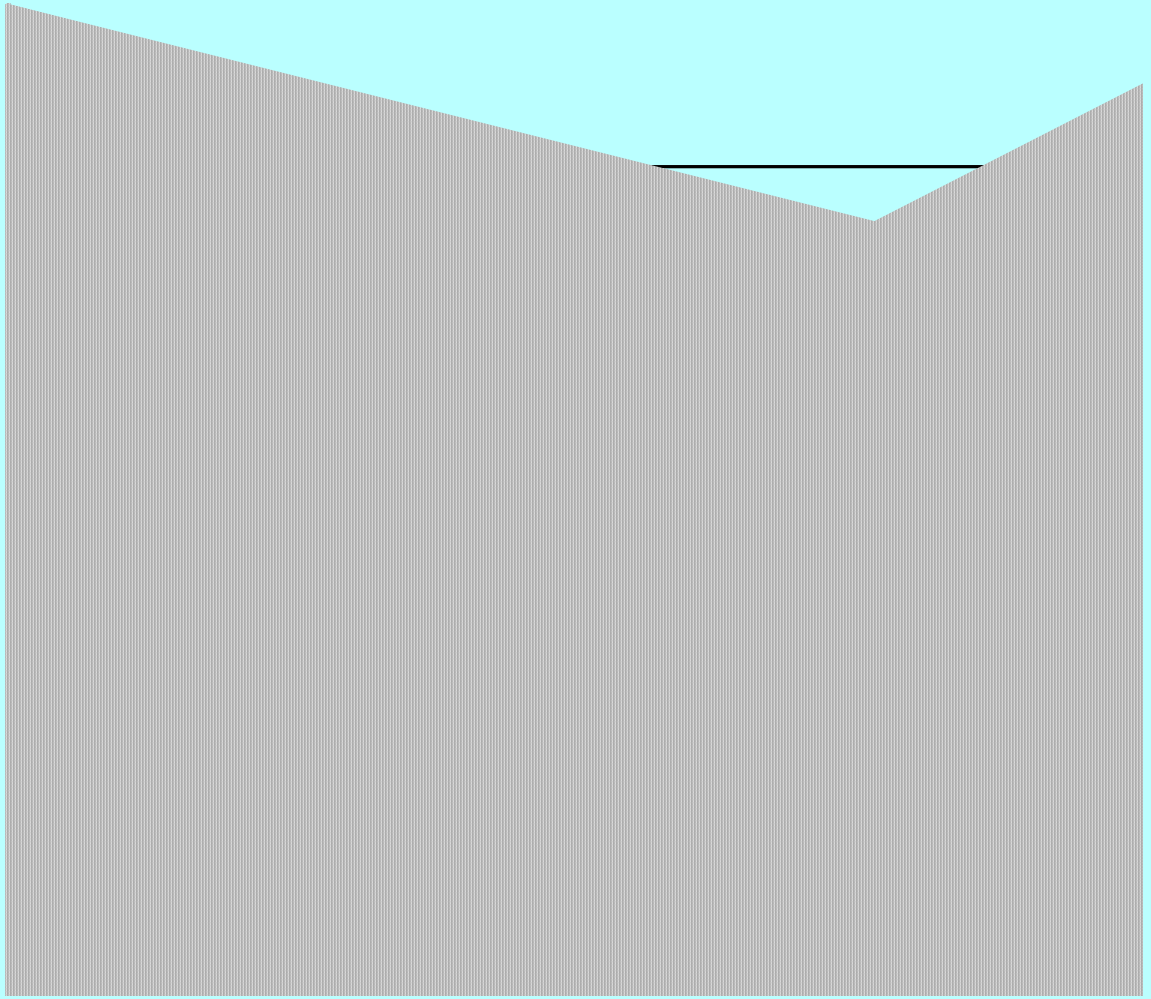
emerging functionalities, unconscious cooperation, collective unaware intelligence must exist also among cognitive agents! (ex. Adam Smith's "invisible hand").

Agent-based SS and

the micro-macro link

the main contribution of AI (and in particular of cognitive agent modeling and MAS) entering the social simulation domain will be an significant advance in the theory of the micro-macro link

Only such a "mind-based" social simulation will allow us to observe at the same time the mind of the individual agents (beliefs, desires, decisions) and the emerging collective action and equilibrium which *co-evolve*, determining each other.



Only MAS can fully deal with this problem

7

Social Functions and Cognition

a) *no theory of social functions is possible* and tenable without clearly solving this problem;

b)*without a theory of emerging functions among cognitive agents social behavior cannot be fully explained*.

Functions install and maintain themselves *parasitical* to cognition:

functions install and maintain themselves thanks to and through agents' mental representations but not <u>as</u> mental representations: *i.e. without being known or at least intended.*

Social Functions and Cognition

While **Social Norms** emergence and functioning require also a "**cognitive emergence**",

Social Functions require an extra-cognitive emergence and working

> *For a Social Norm to work as a Social Norm and be fully effective, agents should understand it as a Social Norm.*

On the contrary the effectiveness of a Social Function is ***independent of agents' understanding*** *of this function* of their own behavior:

a) the function can rise and maintain itself without the awareness of the agents;

b) if the agents intend the results of their behavior, these would no more be mere "social functions" of their behavior, but just "intentions".

“THE core theoretical problem of the whole social science” (Hayek)

"This problem (the spontaneous emergence of an unintentional social order and institutions) is in no way specific of the economic science.... it doubtless is THE core theoretical problem of the whole social science" (von Hayek, *Knowledge, Market, Planning*)

the problem is not simply how a given *equilibrium* or *coherence* is achieved and some stable *order* emerges

Is this emergence just an epi-phenomenon? Is this "order" only from the observer’s point of view?

To have a "social order" or an "institution", spontaneous emergence and equilibrium are not enough. They must be "functional".

Adam Smith's "invisible hand"

Adam Smith's original formulation of "THE problem" is much deeper and clearer

The great question is how:

> **"(the individual) - that does neither, in general, intend to pursue the *public interest*, nor is aware of the fact that he is pursuing it,... is conduced by an invisible hand to *pursue an end* that is not among his *intentions*" (Smith,).**

Hayek like Smith in acknowledging the ***teleological nature*** of the invisible hand and of spontaneous order, cannot avoid attributing to it

> **a (positive) value judgment, a providential, benevolent, optimistic vision** of this process of self-organization (**ideologism**).

In the “Invisible Hand”:

1) there are intentions and intentional behavior

2) some unintended and unaware (long term or complex) effect emerges from this behavior

3) but it is not just an effect, it is an end we “pursue”, i.e. its orients and controls -in some way- our behavior: we "necessarily operate for" that result (Smith).

- **how is it possible that we *pursue* something that is not an intention of ours; that the behavior of an intentional and planning agent be goal-oriented, finalistic (‘*end*’), without being intentional;**

- **in which sense the unintentional effect of our behavior is an "end”??**

Theory of “Function”

This problem appeared in other social sciences as the problem of the notion of "functions" (social and biological) impinging on the behaviour of anticipatory and intentional agents, and of their relations with their "intentions".

The same problems that troubled the theory of functions appear in Smith's theory and in Hayek's view of *social order*.

- the view of the society or group as an organic "order",

- the "positive" view of functions relative to this order.

The problem:

Emergence and Functions should not be
***what the observer likes or notices*,**
("just in the eye of the beholder")

but should be indeed observer-independent,
based on **self-organizing** and **self-reproducing** phenomena,
>>> "positive"- "good" can just consists in this.

We cannot exclude "**negative functions**" (Merton) (**kako-functions**) from the theory: perhaps the same mechanisms are responsible for both positive and negative functions.

- ***Two types of finalistic notions*:**

  - evolutionary finalities, adaptive goals; and
  - mental ends (motives, purposes, intentions).

Intentional behaviour *Vs.* functional behaviour

Finalistic systems: Goal-oriented vs. Goal-governed

There are two basic types of system having a finalistic (teleonomic) behaviour:

Goal-oriented systems - (Mc Farland, 1983),

Goal-governed systems

a specific type of Goal-oriented system based on representations that anticipate the results

- Two levels of functionality: relative to an OverSystem, and *per se*.

"***absolute functions***" those functions that are just emerging and self-maintaining in some MA situation or system, but that are **not "functional to", useful for**, and reproduced by some larger inclusive OverSystem

*"**relative functions***" those functions that are ***functional within* an OverSystems** and ***for*** its goals and its working.

• ***How is it possible that a system which act intentionally and on the basis of the evaluation of the effects relative to its internal goals reproduces bad habits <u>thanks to</u> their bad effects?***

• **If a behavior is reproduced *thanks to* its good effects, that are good relatively to the goals of the agent (individual or collective) who reproduces them by acting *intentionally*, there is no room for "functions”** (Elster).

If the agent appreciates the goodness of these effects and the action is replied in order to reproduce these effects, they are simply "intended".

>> **??** a **behavioristic reinforcement layer** (van Parijs)

together with

>> **a deliberative layer (controlled by beliefs and goals) ???**

the deliberative layer accounting for intentional actions and effects,

the behavioristic layer (exploiting conditioned or unconditioned reflexes) accounting for merely "functional" behaviors??

Or

>> **'habitus' for roles and functions** (Bourdieu), and

>> **intentions for personal purposes????**

Our problem is indeed that:

✓ ***intentional actions have functions!***

Goals and beliefs of the agents have functions.

COGNITIVE AGENTS & FUNCTIONS

If agents appreciate some advantages and results
or evaluate some danger

If they understand causes and/or conditions of these effects
they will intend to control these effects
to produce them on purpose
or to avoid them

This principle is unavoidable for both rationality, planning, and cognitive learning.

To account for functions, we should admit *some mechanism that reproduces the intentional action thanks to (some of) its* effects, but *bypassing agent understanding and planning these effects* (that can even be good for its goals and reproduced for that).

Hayek

> does not adequately explain *for whom* the emerging order is good and to what extent power differences are involved in maintaining it;
> he does not analyse the problem of the effects of our actions that are harmful to others and not to ourselves;
> he does not include in the theory the possibility that the social agents do not know what is best for them;
> he neglects the fact that desires and preferences are not a given (with respect to which the order is good) but are a product of the order itself;
> he makes use of models of group selection that are not clear, etc

The central issue in this talk is therefore

whether it is *possible to recognize and account for the teleological, teleonomic, functional character of the 'invisible hand' without having to adopt a teleological and providential view of society and of history.*

Hayek

The fundamental problem is

how to graft *teleological* but *unintentional* behaviours precisely on *intention-driven* behaviours.

What answer can be given to **Elster** according to whom the idea of *intention* makes that of the *function* of behaviour impracticable and superfluous.

How can intentional acts also be functional, that is,
unwitting but

***reproduced precisely as a result of their unintentional effects*.**

Why also kako-functions?

- the mechanism that install a bad function can be exactly the same installing a good one

- to definitely separate a functional view of behavior and society from any teleological, providential view (functions can be very bad and persist although bad)

- kako-functions cannot be explained in a strictly behavioristic framework of reinforcement learning: the result of the behavior can be disagreeable or useless, but the behavior will be "reinforced", consolidated and reproduced.

20

Hayek's *good order*

He is right in characterizing the long process of the emergence of social order and the formation of institutions, in terms of:

- **"adaptation"** and
- **"selection"**

- **too optimistic view of such an "order" and of evolution**
- **pan-selectionist**

> ***The evolution of the society selects and records the positive results of the experience, after innumerable trials and errors. Only the positive features and the "right rules" survive. The behaviours encouraging the development of the group are persistent and are replaced only when more efficient behaviours have been developed. All behaviours proving antithetical to the group cannot persist and are eliminated*** (Hayek, 1973).

What emerges is not any order

but a good or even the best possible order

Hayek's *good order*

"these **self-persisting** wholes/social structures represent a *conditio sine qua non* ***for the attainment of many individual aspirations***: they go to make up the environment that makes possible the conception of our ***individual desires*** and allows them to be satisfied. ....The emerging social structures are **'useful'** as they form the premises for future human development" (AR VIII).

These institutions ***"are a necessary condition for the attainment of conscious human purposes "*** (AR VII note 5).

>> Very poor model of intentional action and cognition

Hayek’s *good order*

- in the first place, social order - although unconscious and unintentional - is functional, is good/useful only as a function of **the individuals' conscious purposes**.
  In fact, there is only one valid teleonomic notion: the psychological one!
  There can be no *autonomous* teleonomy deriving from an evolutionary perspective.

- Secondly, it is assumed that these emerging social structures are self-persisting precisely *because* they are ‘useful', *because* they realize the **individuals' desires** and their conscious purposes.

Hayek's good order

If the agents fight unremittingly for their own individual good an emerging equilibrium cannot be contrary to their good.
If the emerging good was not good for the agents, it would not stabilize, it would not be maintained, the agents would (consciously or unconsciously) rebel against it, they would react in the direction of their own good.

- Not only does a natural, unplanned order exist, but this order is also useful for individuals, and perhaps the best available among the possible variants

- If it was possible to achieve a better equilibrium for and among these agents, they would find it; if more suitable solutions emerged they would be handed down

Hayek's *fallacies*

1. Whose human purposes?

when we say that the emerging order is "*at the service of important human purposes*" or that it is a necessary condition for attaining conscious human purposes" we must immediately ask ourselves "the human purposes of **which individuals**?!", "**good for whom?!**".
It is not possible to overlook this in an individualistic framework. Do purposes exist for mankind?

even if it were true that this order were not only natural but also a good as far as the agents are concerned (and I will deny that this is really true), *it would be a good for some agents, but not for all.*

For ex. **Distributive?**

Hayek's *fallacies*

2. Ignorance of good

Hayek's view of intentional action is highly limited and inadequate. He fails to consider many cognitive preconditions of the action.

He takes these cognitive preconditions for granted as well as the fact that the agents are capable of perceiving and reacting appropriately to what is bad for them.(Elster's *Sour Grapes*).

> *such an order might not be maintained through its goodness-usefulness but rather as a result of its opaqueness and the strong constraints it imposes on the local conditions and individual choices.*

3. Domination and differences of Power

- Different power of individuals, groups, and clasess
- Unplanned, non-designed, self-stabilizing orders difficult to influence and change.

Hayek's fallacies

4. Self-produced goals

to the extent to which the stability of social order is due to its being satisfactory, this order produces the needs and the mentality that it must satisfy in order to be stable!

Moreover, *order does not emerge and become stable only because it is satisfactory.*

5. Path-dependency, conventions, and badness of the invisible hand

the convention -although worse than a feasible alternative - is *retained also by its perverse effect of* ***irreversibility***: indeed, it is precisely the fact that it is preferable not to deviate from it *individually* - which is the result of its having become a "convention" - which perpetuates it.

Hayek

Back to the main issue:

Do actions directed towards good
imply the goodness of the emerging order?

from the fact that individual behaviour is directed towards good (individual subjective goals) *it does* ***not*** *follow that the emerging order is good* (for individual subjective goals).

30

Hayek

How can agents who intend their results and prefer what is good for them to what is bad for them not only produce also perverse and harmful side effects, but allow the latter to organize themselves and direct their behaviours.

The invisible hand, spontaneous order, precisely because it is an unintentional and yet **emerging** function, **self-organizing** and **self-reproducing** (through the individual behaviours),

is substantially indifferent to the goals and good of individuals,

and may be addressed equally to good and to evil (Leopardi's philosophic view is here opposed to Hayek's philosophy).

Hayek's optimism is theoretically unjustified

Unexpected evil effects exist, or evil effects combined with good individual intentions (Boudon, 1977) in which

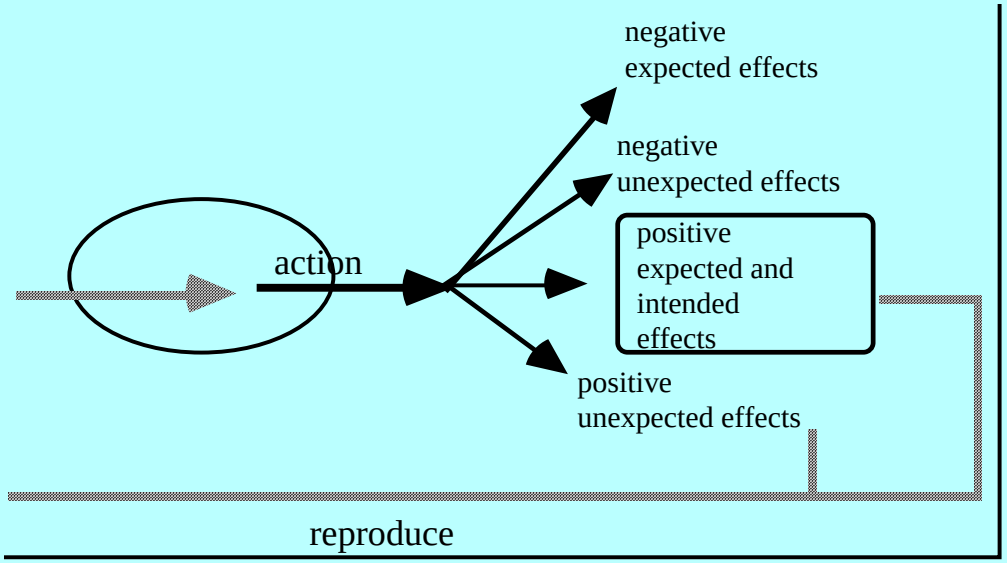
the intended good effects reproduced

***in spite of* the negative consequences**.

This is true,

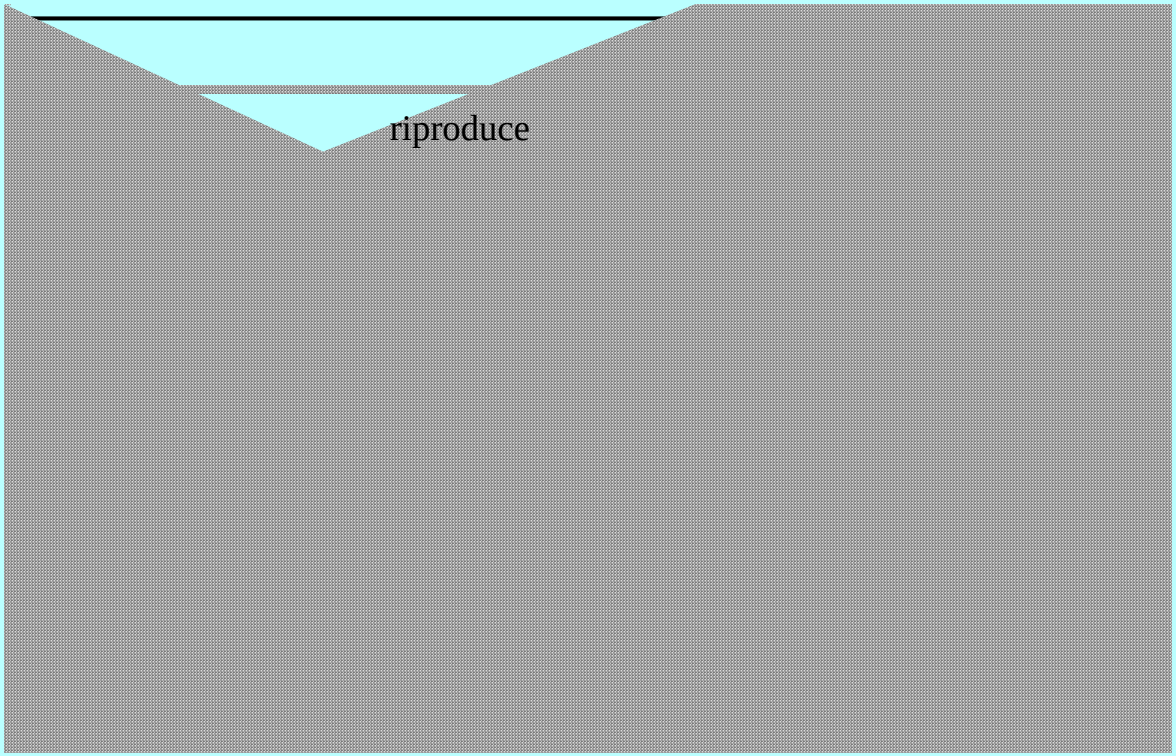
- both in the case in which the **evil effects are not perceived** or are not attributed correctly,
- and in the case in which they are **perceived**

(in the second case the good effects must be subjectively more important and in any case preferred (for instance, be closer in time), or else are more conditioning/reinforcing than the evil effects)



But there are also

harmful effects capable of self-reproduction (through the action) precisely *because of their negative nature* (Castelfranchi, 1997; 1998b; 1998d).



a long line of automobiles and the slowing down due to the simple individual intention of rapidly glancing at an accident that has occurred in the other lane

The notion of ‘function'

as an effect selecting and reproducing its own cause

How is it possible for a system that acts intentionally on the basis of an evaluation of the effects vis-à-vis its own goals, to reproduce bad habits precisely *as a result of* their bad effects?

And even more crucially - if a behaviour is instead reproduced thanks to its good effects with respect to the (individual or collective) goals of the agent who reproduces them by acting intentionally, then there is no room for the "functions".

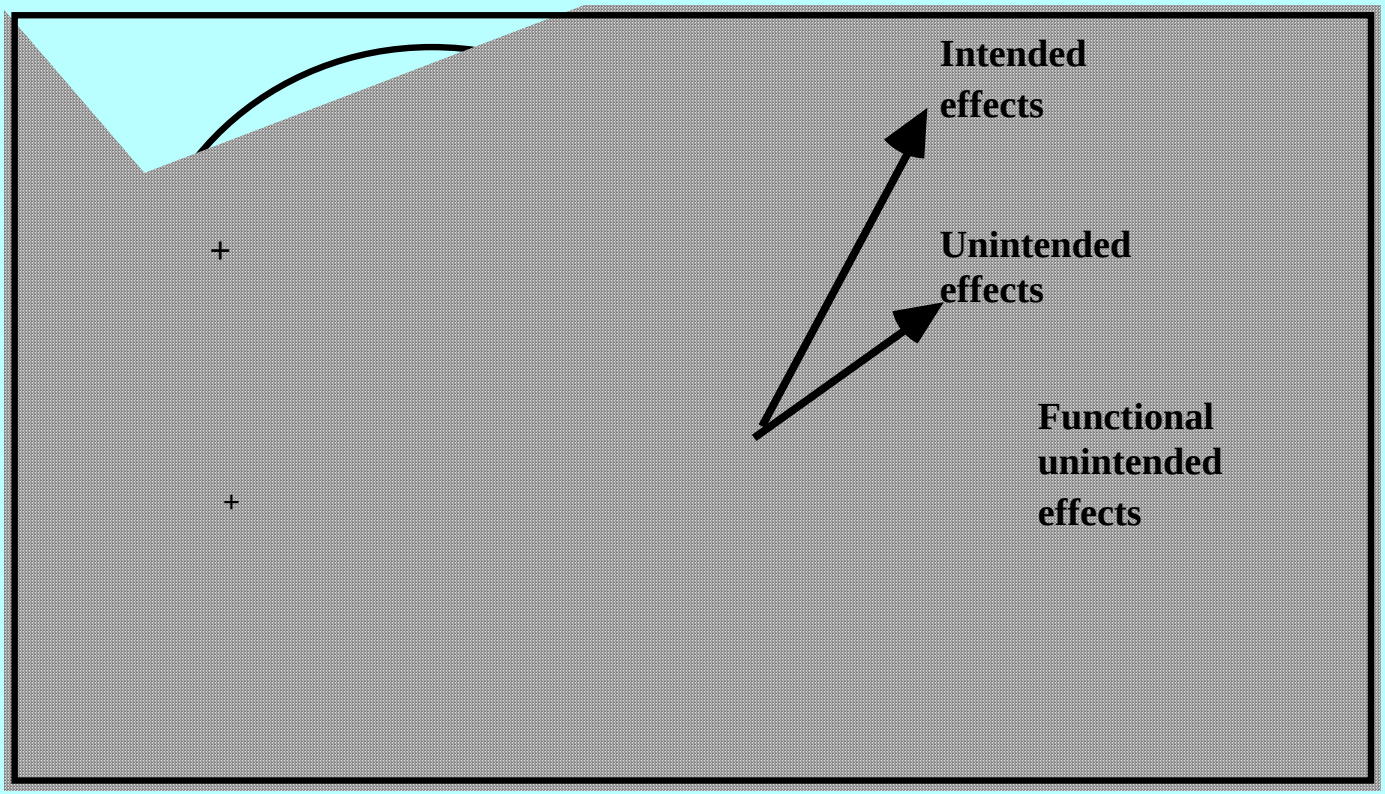
It is necessary to have **complex reinforcement learning forms** not merely based on classifiers, rules, associations, motor sequences, etc. but ***operating on the cognitive representations governing the action, that is, on beliefs and goals***.

In this view *"the consequences of the action, which may be more or less consciously anticipated, nevertheless modify the probability of the action being repeated the next time in similar stimulus conditions "* (Macy, 1998). More exactly:

> ***the functions are simply effects of behaviour which go beyond the intended effects but which can successfully be reproduced because they reinforce the agent's beliefs and goals that give rise to this behaviour.***

How Social Functions are <u>implemented through</u> cognitive representations

The basic model



There are two *Cognitive "reinforcement" principles*:

1. Belief Reinforcement:

two different mechanisms can be postulated:

association (accessibility) :

the *association* between the belief and that context or scenario is strengthened: the believe will have more probability to be retrieved next time in similar situations; it will be more activated, more available and accessible (accessibility bias);

confirmation (reliability) :

some of the action's effects are perceived by the agent (even if not necessarily understood and causally connected to its actions) and they *confirm* the beliefs supporting the action: they give new *evidence* for that belief, increase its "credibility", and reliability: they augment its “truth” or the subjective probability of the event.

2. Goal Reinforcement

two different mechanisms can be postulated (analogous to the beliefs reinforcement mechanisms):

association (accessibility) :

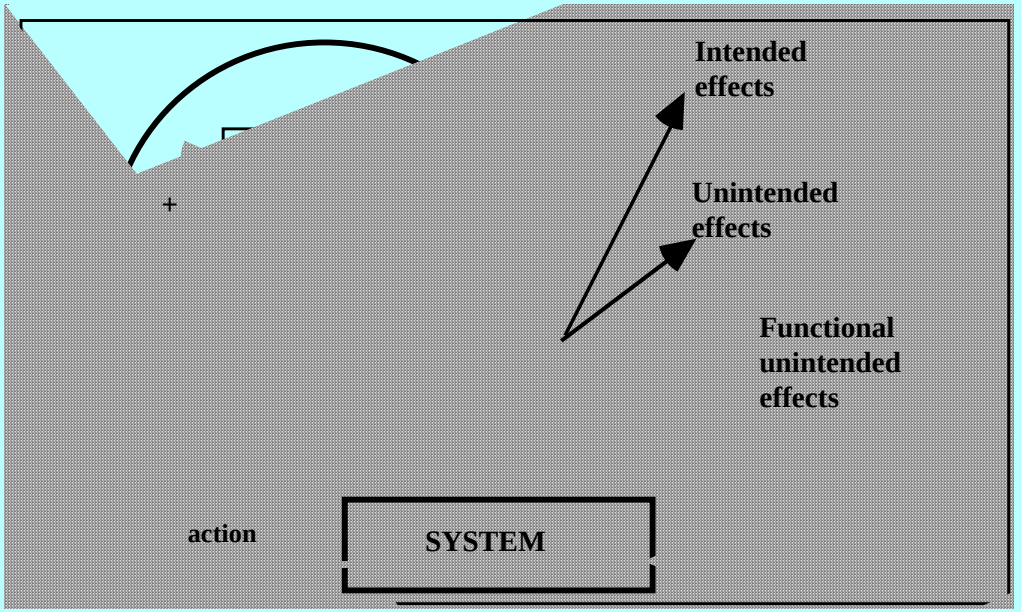
the success of the chosen goal, plan, action is memorized in the sense that the association between the goal-plan and that problematic context or scenario is strengthened: the goal/plan (solution) will have more probability to be retrieved next time in similar situations; it will be more activated, more available and accessible;

confirmation (reliability) :

the success of the chosen goal, plan, action is memorized; it increments a "successfulness index" relative to that choice; or better some meta-cognitive evaluation of the value of the action. This memorized behavioral choice is "confirmed": next time the probability to choose the same way (goal, plan, strategy, action) will be greater: it will be more preferable and reliable (we will trust more it).

The reinforcement of both the *belief* and the *goal/plan*
will determine *a reinforcement of that behavior*

The feedback reinforcing mechanism from the global level back to the individual can be not only due to a merely combined effects: this feedback can be an action of the OverSystem (Institution) aimed at controlling the individual behavior.



There should **always** be a "closure", a **way down** (from the global or System level to the individual mind) but in this case the individual behaviors (beliefs, goals) that are useful to the system (and reproduce it) are -thanks to their understood effects- reinforced and reproduced by the system:

- *prescriptions*
- *socialization*

Car Crossing

The timid mind

The timid believes

(**B1**) that there is a certain amount of "aggressive drivers" (Class X) who might not respect traffic rules, and that would try to cross a crossing even without having the precedence. This belief has a certain strength or probability in his mind.

(**B2**) that to be careful is better, and that to slow down (and in case to stop) and letting the other cross is careful.

He also has a goal (**G1**) of being careful and of letting the other cross if she is really trying to do so. Its character consists exactly in this belief and in the consequent preference to "let the other win".

He also believes (**B3**), in this specific crossing situation, that the coming driver probably is an "aggressive one" (since she is not slowing down sufficiently): she is a member of the Class X. Then he instanciates the goal: to slow down and (in case) let the other cross. This expectation and this goal induces a careful and hesitating behavior.

The aggressive mind

She believes

(**B1**) that there are several slow, hesitating, uncertain drivers (Class Y) that waste our time. (She could also believe that Norms themselves are stupid things and waste our time).

(**B2**) that if one tries to cross -not slowing down- one succeeds to cross because the other will give up. Thus she has the goal (**G1**) to attempt to cross anyway.

Arriving at the crossing, she does not slow down in time, and, observing the careful behavior of the other (she herself is determining), she will assume (**B3**) that the other is a member of the class Y, and that he will not compete. Thus she will have the goal of not stopping and crossing

Car Crossing

The result is that the bad driver will cross before the other, but there will be no accident.

The most important effect is that:

> ***the expectations*** **of both the drivers relative to the behavior of the other, and to the success of their own behavior (respectively: to pass without wasting time; to avoid accident)** ***are confirmed*****!!**

Basic and general **beliefs** are confirmed.

Both agents, without (necessarily) understanding this, and without wanting this,

> ***produce through their behavior the effect of reinforcing their own and the other's behavior: beliefs and preferences.***

The agents are unconsciously cooperating to reinforce their own and the other's behavior**; in this way the social phenomenon stabilizes, reproduces, is stronger, and** ***spreads*** **around.**

An example: **dirty and clean streets**

A social **(kako)function** based on *social conformity and imitation*.

The agent assumes (**B1**) that this is a bad behavior or even a forbidden one; he assumes (**B2**) that a lot of other people behave this way; that (**B3**) this can be quite practical and easy sometimes; he assumes that (**B4**) his contribution to the garbage is quite marginal and small (that its true). He has the goal (**G1**) to do as others do and until others do so (Bicchieri, 1989); or at least, to do as others do and until others do so if this is useful and practical for his goals. Goal G1 on the bases of beliefs B2, B3, B4 will generate a goal (**G2**) to leave small garbage in the street, which overcomes the possible goal (G3) -based on B1- of not dirtying the city. Now the result of such a behavior is that streets are dirtier; this is perceived and then it will confirm the supporting beliefs (B2, B4) and the goal G2.

Everybody reinforces the behavior of the others.

The global effect is not wanted and intended by anybody; **the reinforcement effect is also unattended and unintended.**

> **The behavior is (reciprocally) reinforced by its effects. <u>These effects are self-maintaining and reproducing through the reinforcement of their own causes</u>. This *passes through the mind of the agents* (their beliefs and goals) but not through their consciousness and intention.**

It is quite interesting to observe that exactly the same kind of beliefs, and an identical goal (**G1**) can generate in this case an eu-function: to maintain the city clean.

'Natural and regular evil' (Leopardi versus Hayek): the stability of negative effects and evil-oriented functions (kako-functions)

It is of decisive importance **to realize that** ***negative effects can stabilize and self-reproduce like positive effects*****.**

The only differences are that:

a) Positive and negative effects are not equiprobable, since behaviour, intelligence and learning are oriented towards the production and preservation of the positive ones.

b) It is slightly more likely that the negative effects will be discovered and specifically resisted

Nevertheless these differences do not eliminate the fact that:

the negative effects can stabilize and self-reproduce exactly like the positive effects.

Individual level: *vicious circles*

The wrong treatment and the persistence of pain

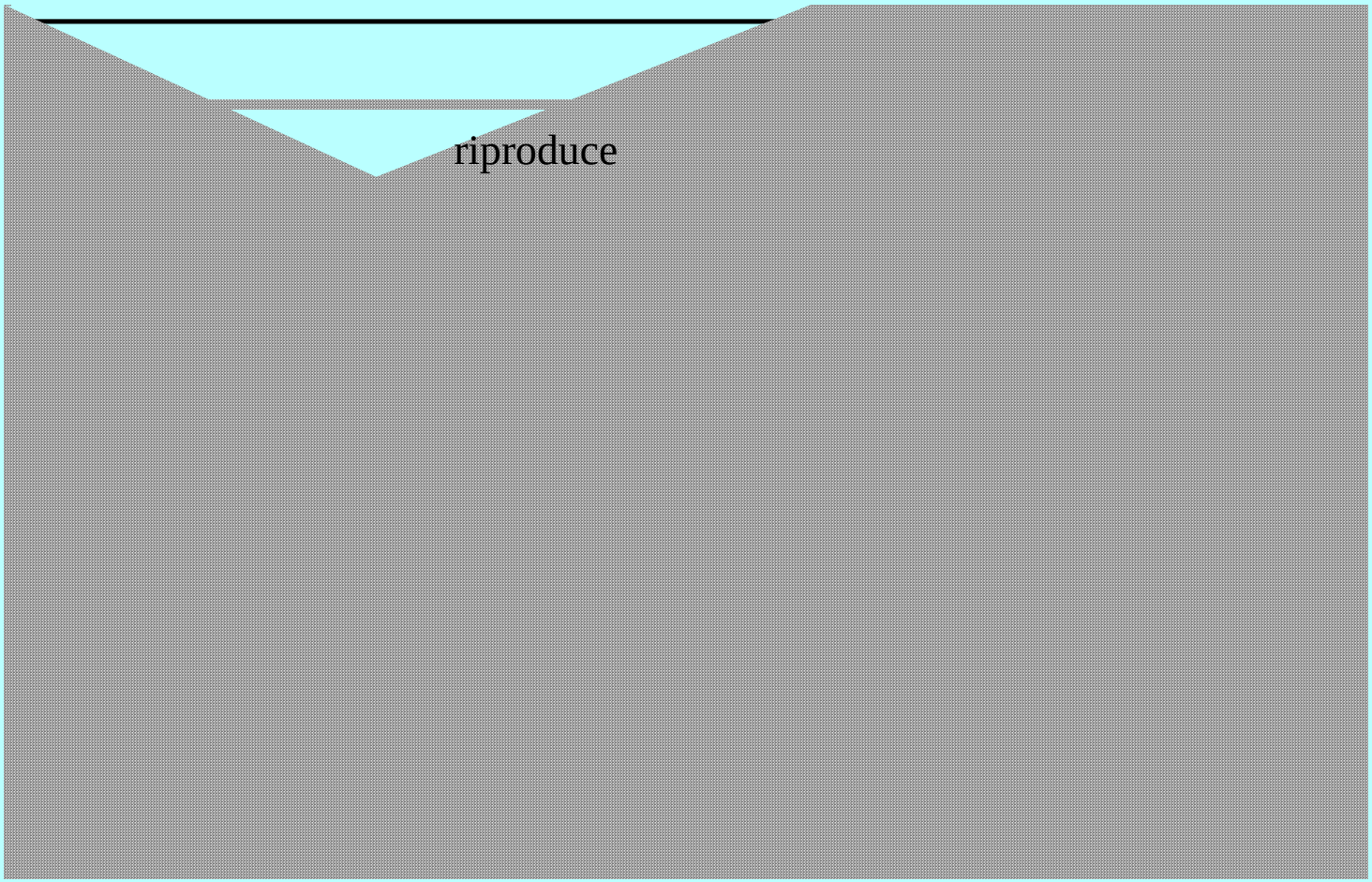
Superstition etc.: an anomalous reinforcement

✓ *The anxious mother*

On the one hand, the negative effect - although anticipated - is not an "intention" of the mother, it does not motivate her behaviour; on the other, *the behaviour exists and is reproduced also thanks to the unintentional and negative effects* therefore *for* them: the behaviour is goal-oriented but not goal-directed towards them**.**

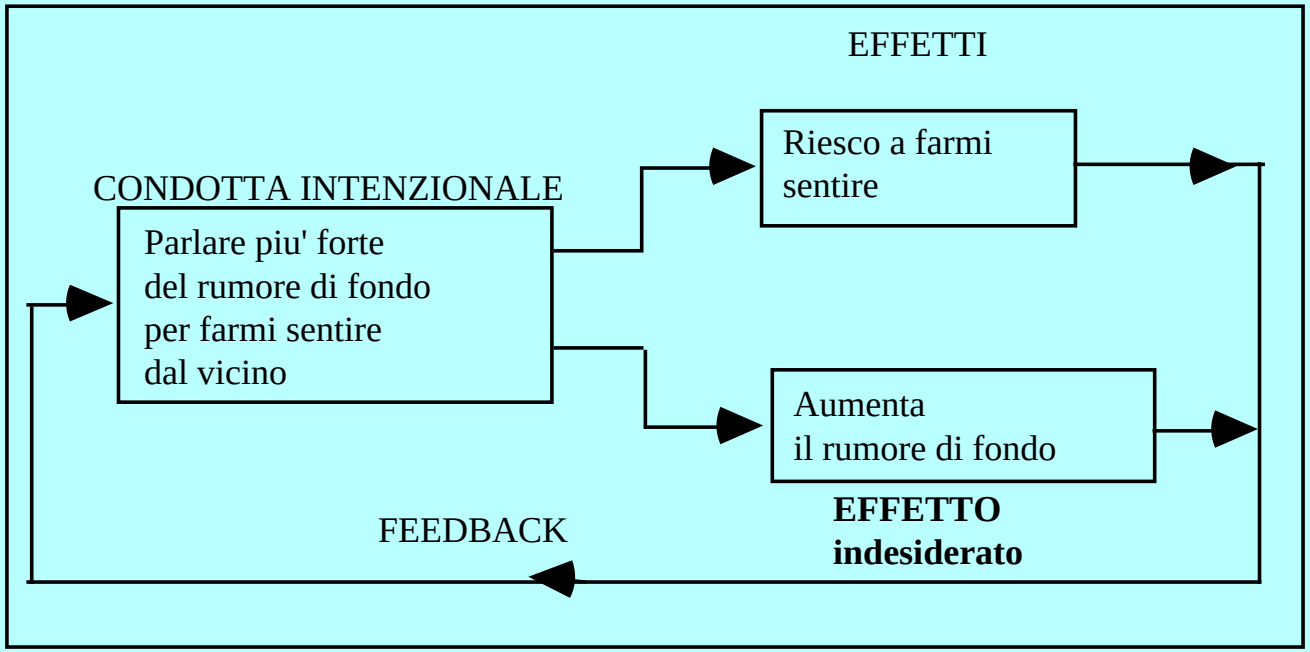
The general abstract model (in the sense that it corresponds to a range of micro-mechanisms) would be as follows:

> ***an intentional action is repeated with a view to its goals although it is actually reproduced either by the (complete/partial) failure or by negative effects (perceived or not perceived; understood or not understood) that repropose the problem and reinforce the beliefs and goals on which the action is based.***


riproduce

Interpersonal and collective level: *vicious circles*

Hubbub in a restaurant or at a party



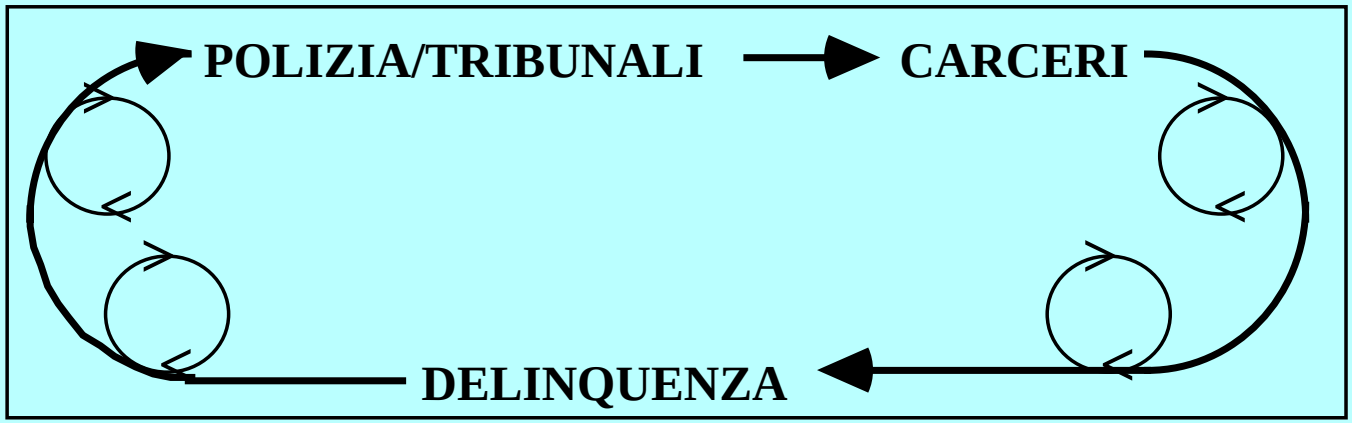
(The example given is merely -on a small scale- the model followed by the **arms race**).

Street litter

Hostility leads to hostility

Institutional level: *vicious circles*

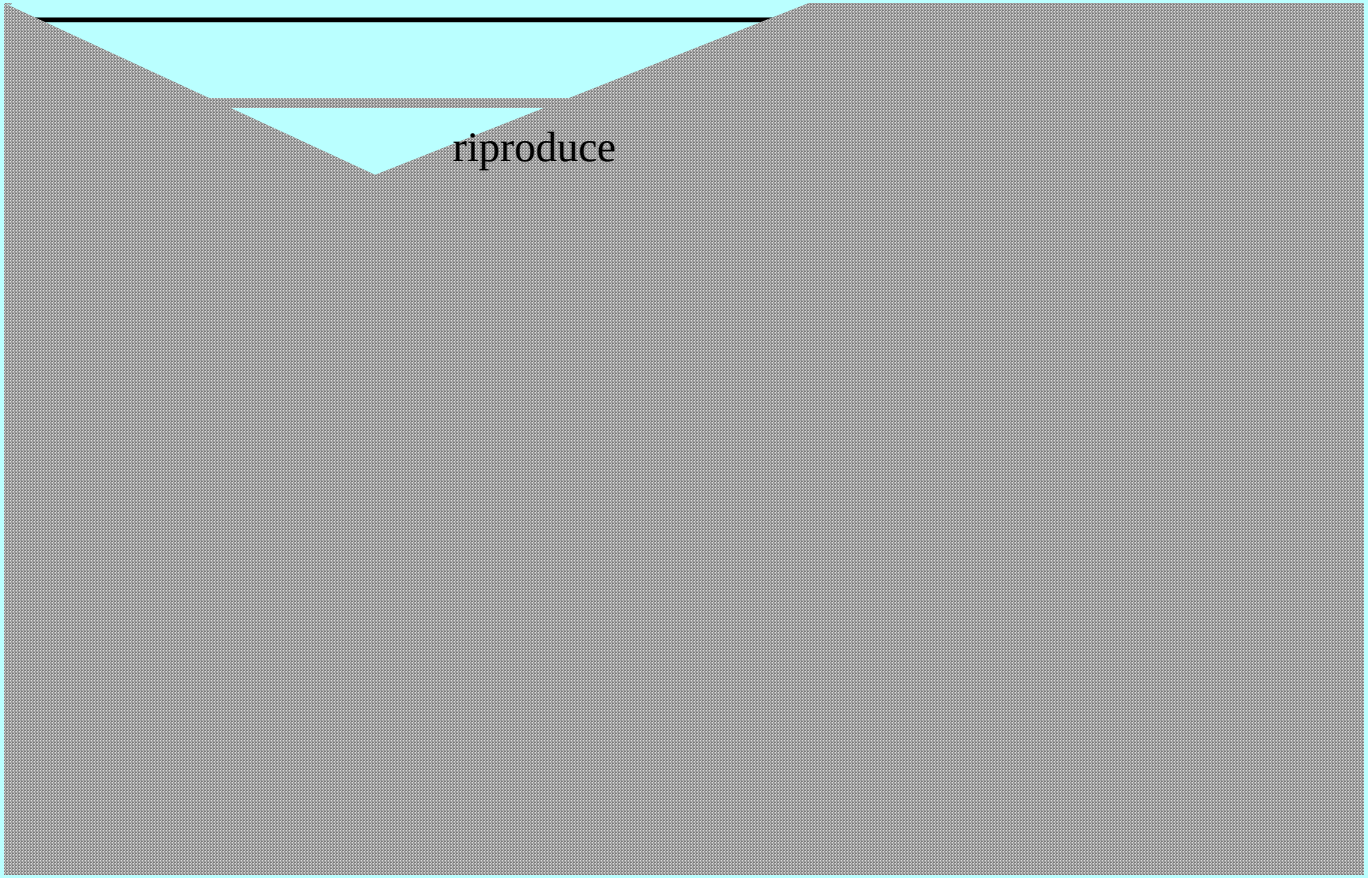
Prisons <==> Delinquency



Vicious circles –

Non intended effects –

kako-Functions



What distinguishes *function* from *non function* is not that the unintentional (collective) effect is *good* but that it is *self-organizing* and *self-producing* by means of positive feedback, that is, by reinforcing, selecting, and reproducing the behaviour that generated it:

***unintended effects that select their own causes*.**

>> *functionality* must be kept distinct from *goodness* (and that is from the subjective goals of the agents),

>> good and bad functions (exactly like unintended good and bad effects) are on the same plane: both may be *self-organizing*.

>> the function is not reproduced or maintained or repeated by virtue of its good effects (a risky approach owing to the boundary with intention)

the psychologistic and subjectivistic *reduction* of functions to what is 'good' for the agents
carried out both by Hayek and Boudon
actually

- amounts to a ***liquidation of the notion of function***

- and completely blurs the philosophers' intuition that there is a form of autonomous ***'end'*** at this level of *organization*.

Conclusions on Hayek

Hayek is **definitely right**

- when he claims that the invisible hand is the theoretical heart of all social sciences;
- he is right when he claims that the same results could never be achieved through centralized knowledge, decision-making and planning;
- when he claims that many results with positive effects for persons can be achieved unconsciously and emergently;
- and when he uses the notion of 'function', albeit to a limited extent.

The **dissent** is related to

the decline of any critical attitude towards the invisible hand and spontaneous order;
when tautologically each and every order that is stabilized, *provided that it is spontaneous*, becomes good *for the agents*.

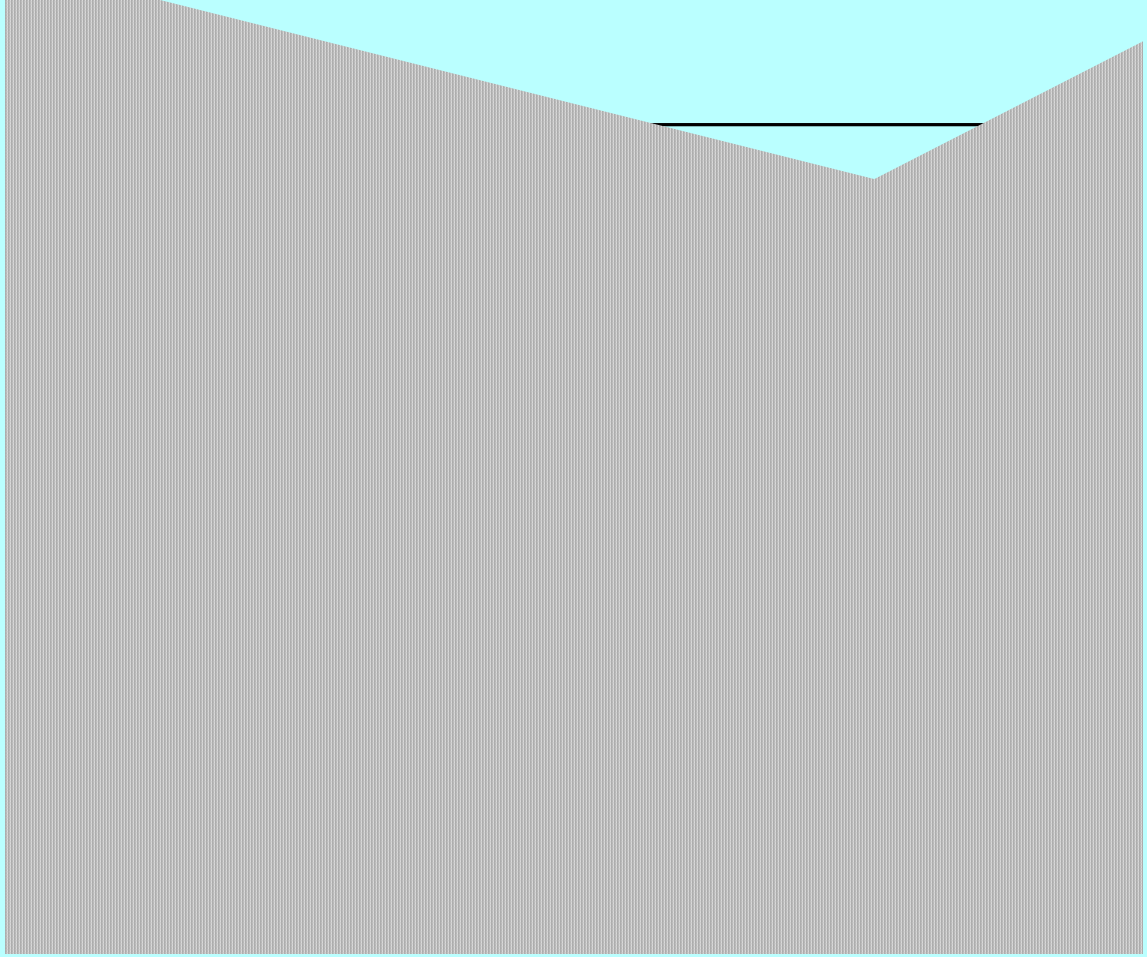
the systematic **confusion** between '**ends**' in a purely self-referential, evolutionary and functional sense, and '**ends**' in a subject sense.

I have tried to argue that these two distinct notions can and must be kept separate and each has an autonomous foundation, and that they are each scientifically feasible. But - as I have said- one unpleasant consequence of this is that

self-referential teleonomic phenomena (such as spontaneous order, social functions, conventions, etc.) are not guaranteed to be functional to human needs, to be good for subjective human purposes.

We would have to have a much less providential conception of the invisible hand and a much less optimistic one of spontaneous order.

The invisible hand, like everything 'natural', is by nature *indifferent* to the good and the interests of individuals (Leopardi).



Only MAS can fully deal with this problem

>> Up & Down

The feedback problem:

- **Which is the information locally needed for individual adjustment, necessary for producing the resulting 'equilibrium' or 'order' or 'desired structure', and coming back from this emerging structure.**
- **Is it a *local* information or a *global* one?**

von Hayek (and others): the **price** as the necessary local *information* about the global dynamics between supply and demand; local feedback from and for the global dynamics.

Ex. in a line an approximate line (linear) structure (quite global information) and the position of the last guy (the one after whom I have to locate myself) (local information but determined by the global structure)..

The ‘END’

(in a third sense)

An Emergent Confusion

bad conceptual and epistemological state of the triumphant notion of "**emergence**"

> **Diachronic emergence**

> **Synchronic emergence**

Gestalt emergence

- it *does not require "complex" or "chaotic" system.*
- it is very clearly subjective,*"observer relative"*, and this might not be true for other forms of emergence. (Virasoro, 1996)

An Emergent Confusion

Also in SocSim, what has been called an "emergent" solution, be it intelligence, cooperation, or whatever, is frequently enough just a structure the observer-designer tried again and again to build up, and that he founds interesting for his purposes;

no causal effects on the phenomenon itself,

not really self-maintaining and self-reproducing,
acquiring objectivity and independence from the observer-designer.

> Descriptive emergence

Complex systems, consisting of many active elements, can be described either in terms of the actions/properties of their components or at the level of the system as a whole. At this level it may be possible to discover a concise description using new predicates for new properties and regularities which are "emerging" because they are only at the global level.

Main question:

is this notion necessarily relative to an observer and to her "view" and evaluation; is it necessarily subjective; or

is it possible a scientific notion not based on perception, description and interest of the observer but on the self-organization of the phenomenon in itself?

An emergent structure is **objective** when there is some ***specific causal effect on its environment due to the global, structural properties in themselves***;
and it is objective and independent in a stronger sense when *it reproduces itself thanks to these effects* (**circular causality**).

> Cognitive emergence: "*immergence*"

Hayek

How Social Functions are implemented through cognitive representations

First, behavior is goal-directed and reasons-based; i.e. is intentional action. The agent bases its goal-adoption, its preferences and decisions, and its actions on its Beliefs (this is the definition of "cognitive agents").

Second, there is some effect of those actions that is unknown or at least unintended by the agent.

Third, there is circular causality: a feedback loop from those unintended effects to increment, reinforce the Beliefs or the Goals that generated those actions.

Fourth, this "reinforcement" increases the probability that in similar circumstances (activating the same Beliefs and Goals) the agent will produce the same behavior, then "reproducing" those effects.

Fifth, at this point such effects are no more "accidental" or unimportant: although remaining unintended they are teleonomically produced: ***that behavior exists (also) thanks to its unintended effects; it was selected by these effects, and it is functional to them.*** Even if these effects could be negative for the goals or the interested of (some of) the involved agents, their behavior is "goal-oriented" to these effects.

Even accepting the postulate that *individuals tend towards their own happiness or well-being*,

- it does not follow that by so doing they produce their own good, either individual or collective, or objective or subjective.
- It does not follow that the emerging equilibrium and spontaneous order are good.
- Not only can "collective" results be bad - as in the classical prisoner's dilemma (why would spontaneous order be precisely a huge PD with many players? What would guarantee the public good (Smith) or the human purposes?
- Equilibrium may be bad also for the individual subjective goals, as not only can the good effects (for the individual and individuals) be self-organizing; also the negative effects (for the individual and individuals) can be self-sustaining and self-reproducing. It is not simply a matter -as is obvious- of producing also ‘disorder’ (which is combated by vital activity and action), or there being undesirable and harmful effects;

it can emerges precisely
an *order* of harmful effects.

Programming (with) ‘the Invisible Hand’??

The problem of ***Emergent Computation*** and **DAI/MAS**

Central themes of EC include (Forest 90; Todd 93):

- **self-organisation, with no central authority to control the overall flow of computation;**
- **collective phenomena emerging from the interactions of locally-communicating autonomous agents;**
- **global cooperation among agents, to solve common a common goal or share a common resource, being balanced against competition between them to create a more efficient overall system;**
- **learning and adaptation (and autonomous problem solving and negotiation) replacing direct programming for building working systems;**
- **dynamic system behavior taking precedence over traditional AI static data structures.**

Typical DAI, MAS issues **Ulieru**

- **Also higher level components: complex AI agents, cognitive agents**
- **i.e the problems of human society: functions and ‘the invisible hand’**

spontaneous emergence of order, beneficial self-organisation, the impossibility of planning but also harmful self-organising behavior

Order versus happiness: relative goodness and goodness in itself

There are **two different and independent criteria** with reference to which something may be deemed to be good or functional.

Let us call the first criterion “**relative**”. It presupposes a goal-directed (and if possible, goal-regulated) entity X, that is, an agent possessing its own goals. What favours one of X's goals is “good for X"; what damages or threatens one of its goals is “bad for X”.

The second criterion is instead **self-referential or "absolute"**.

In this case "good" or "functional" simply means capable of self-reproduction, self-sustaining through a reproductive cycle and by virtue of a positive feedback within this cycle. There is not necessarily any reference to someone's needs, goals, well-being or happiness.

Hayek

Hayek's *good order*

Smith explicitly states that the resulting equilibrium is in the "public interest" and to the good of the nation, and it is clear that the invisible hand providentially guides us to *unconsciously pursue* the general good. Moreover, he is the apologist of Mandeville, of private vice and public virtue: selfishness produces the general good.

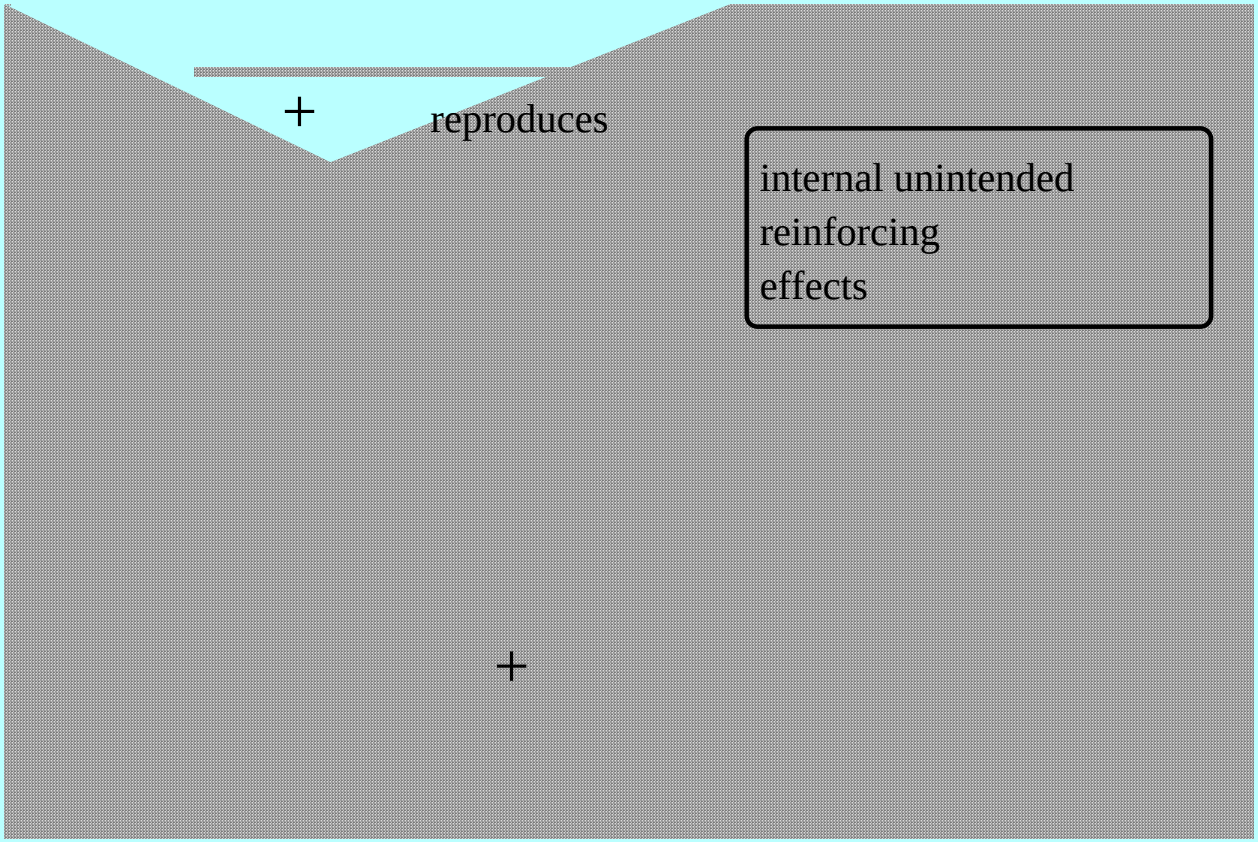
It is more surprising - in view of his sharp mind and critical spirit - to find in Hayek (as it seems to me) exactly the same **fallacy** as in Smith. The emerging "order" is transformed from a mere dynamic equilibrium into something 'good' for mankind (and even something that one cannot seek to improve). It without doubt constantly acquires

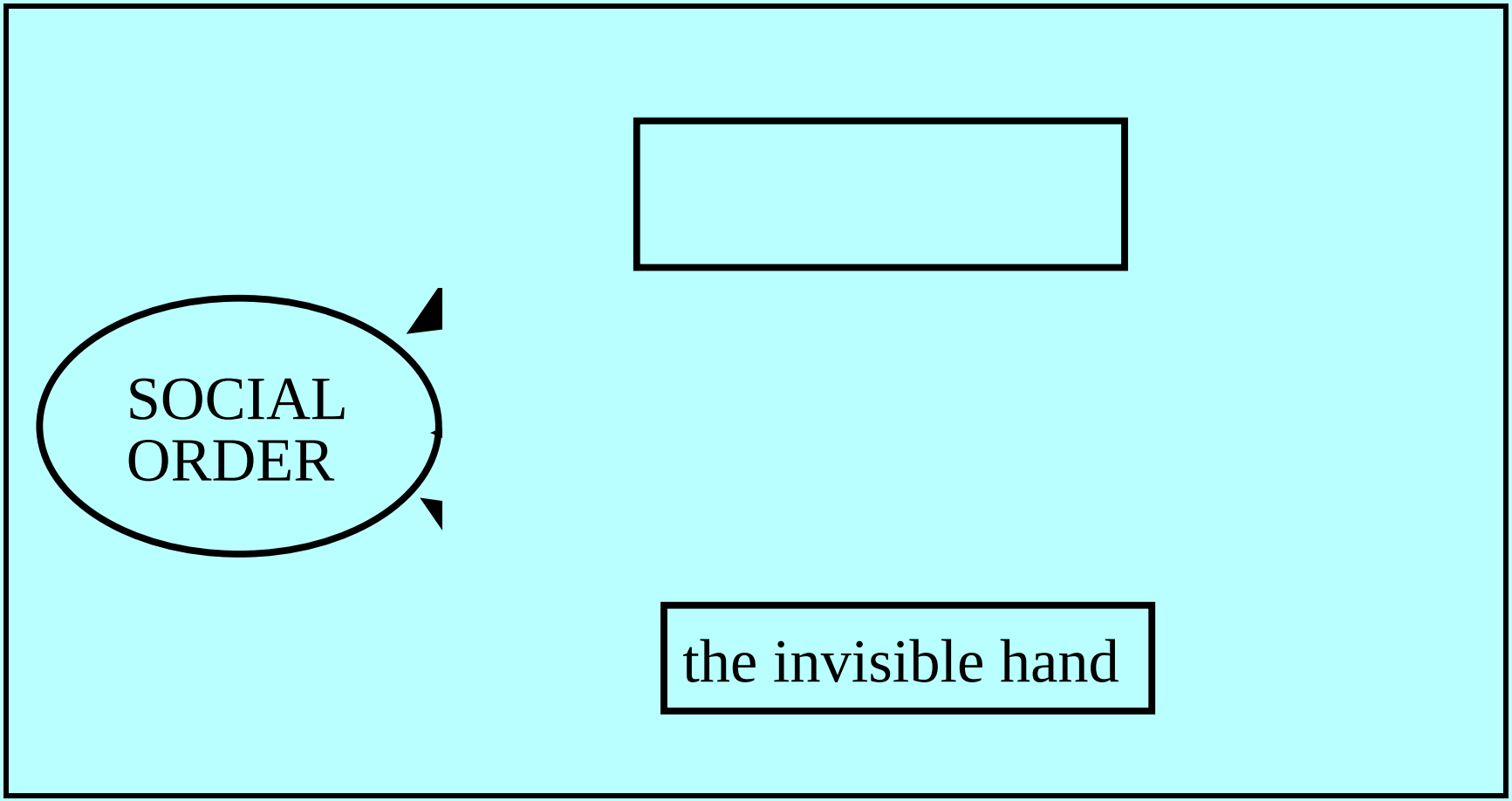
a positive *connotation*

Vicious circles –

Non intended effects –

kako-Functions



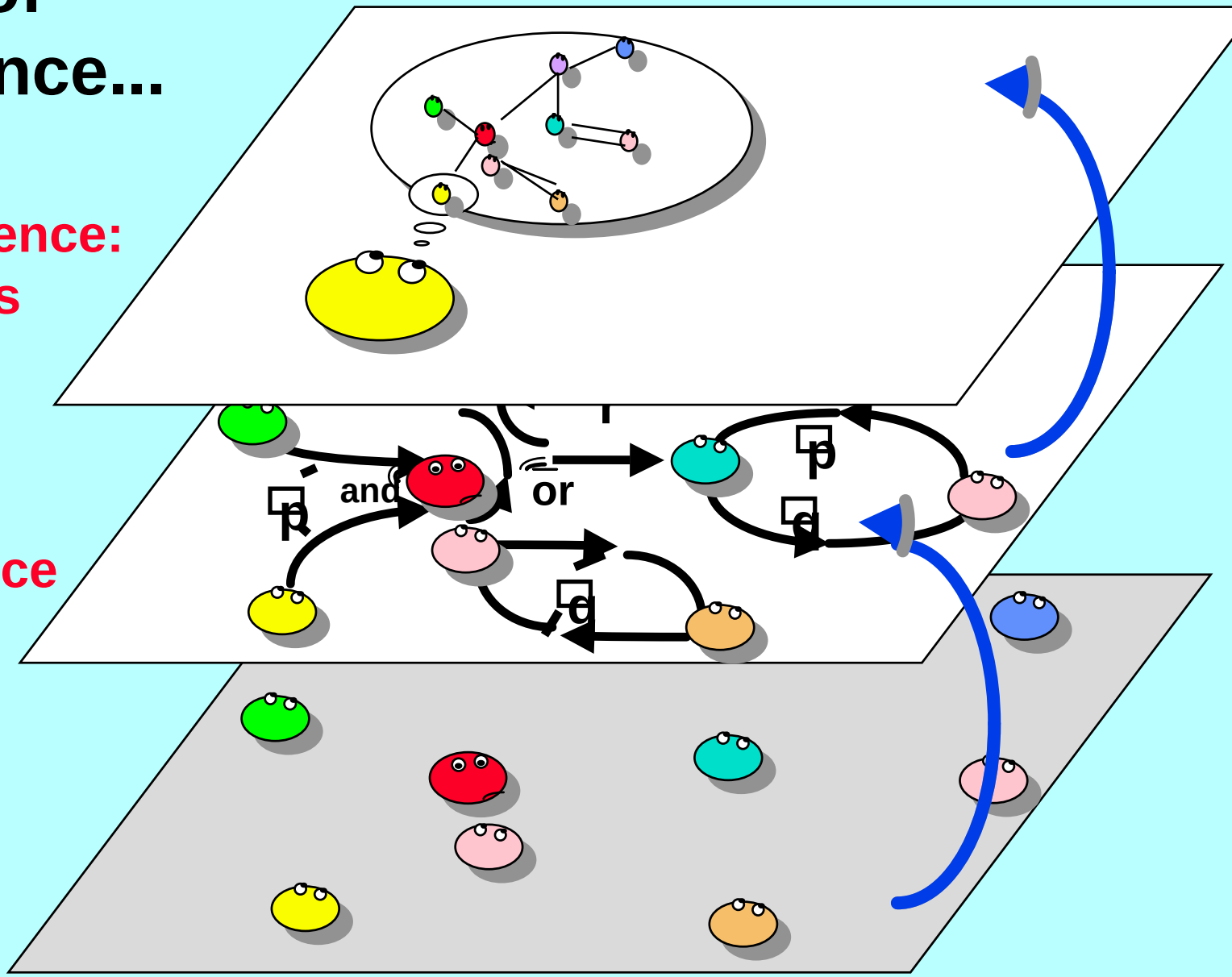

SOCIAL
ORDER
the invisible hand

Levels of emergence...

cognitive emergence: awareness

objective dependence network

agents in a common world

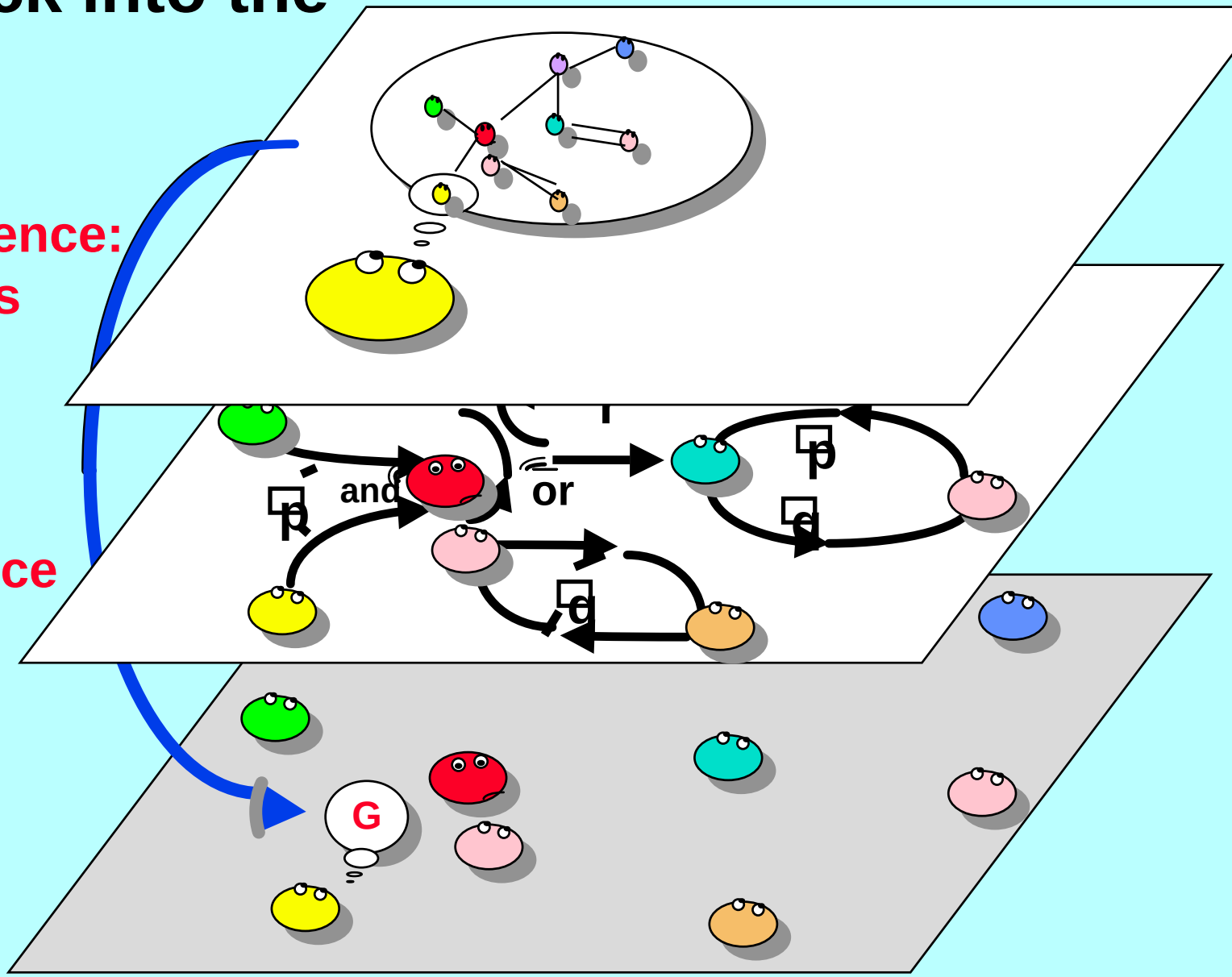


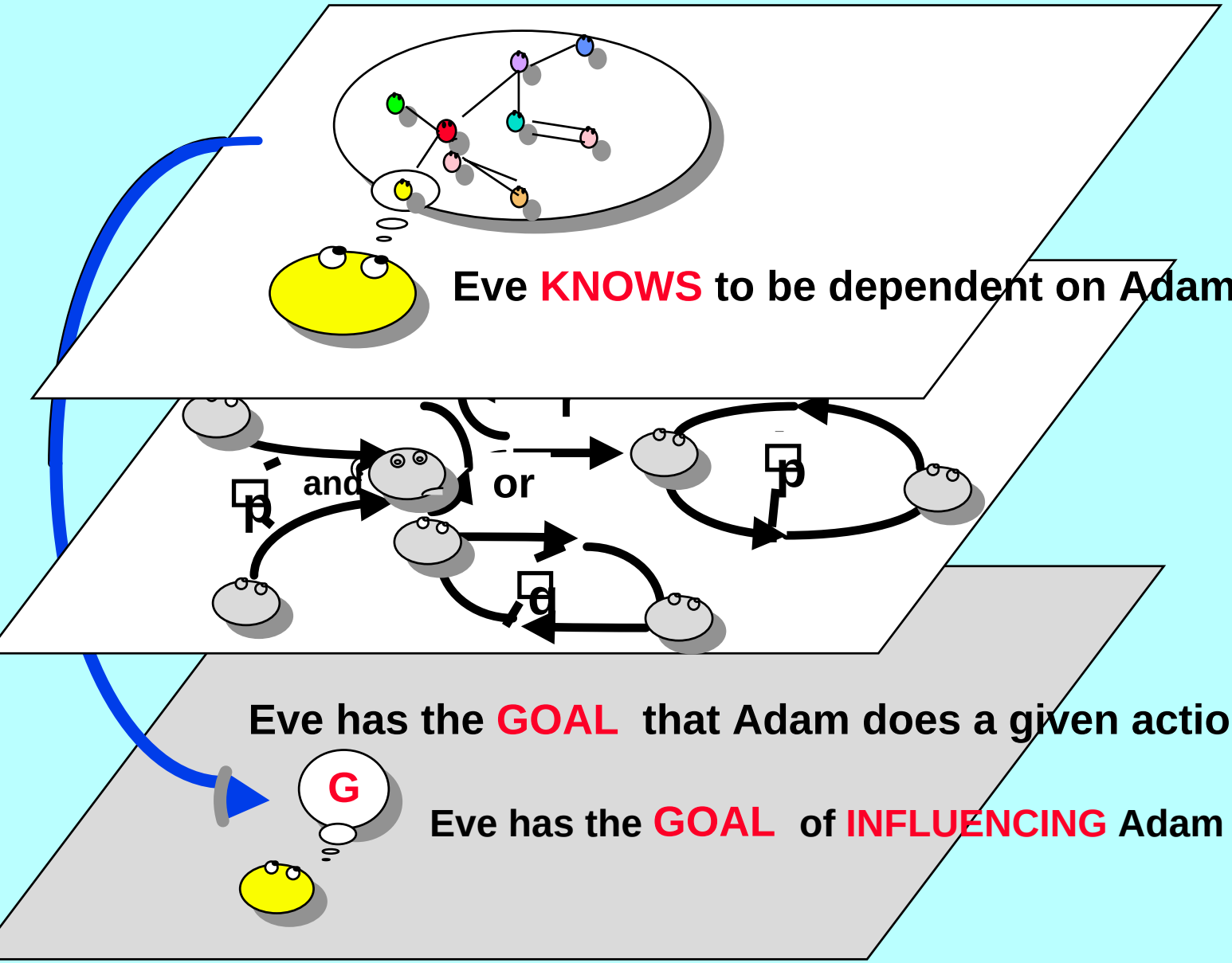
Feedback into the minds

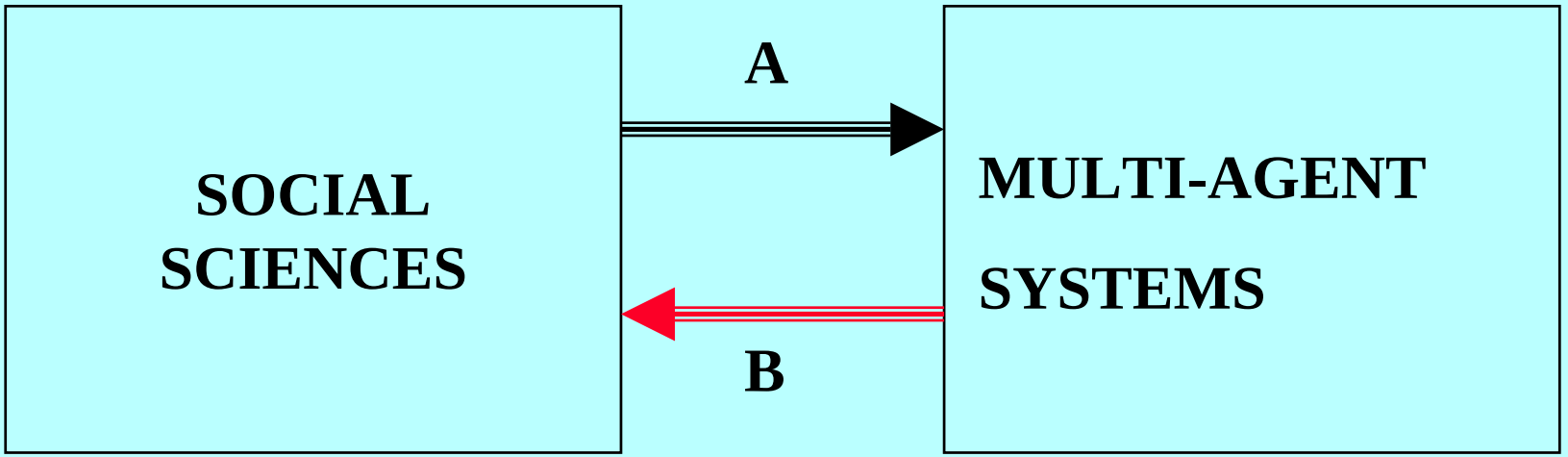
cognitive emergence: awareness

objective dependence network

agents in a common world




Eve KNOWS to be dependent on Adam
and
or
Eve has the GOAL that Adam does a given action
G
Eve has the GOAL of INFLUENCING Adam



B: Not only an 'experimental method' and experimental platforms

CONCEPTS, MODELS, THEORIES

The new COMPUTATIONAL SOCIAL SCIENCES