Between biology and culture: for a cognitive description of altruistic behavior

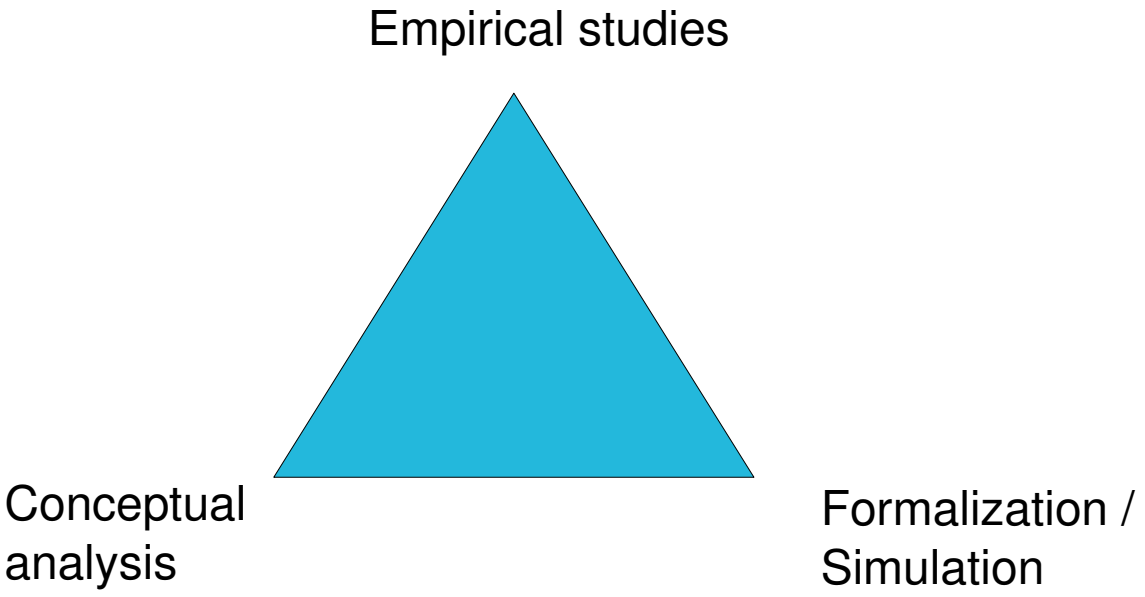
Gennaro Di Tosto, Mario Paolucci, Rosaria Conte
Institute of Cognitive Sciences and Technology, CNR
Rome, Italy

CSS-TW1, Bertinoro, 27 Maggio 2006

Evolution of cooperation

Input: Contribution from different fields of research.

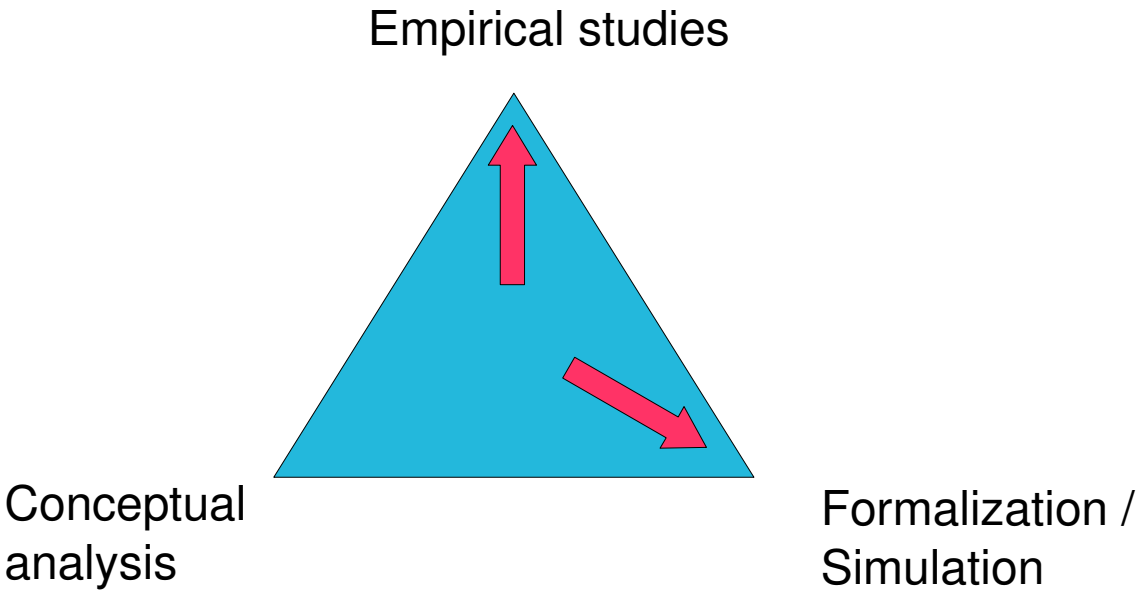
Output:



Evolution of cooperation

Input: Contribution from different fields of research.

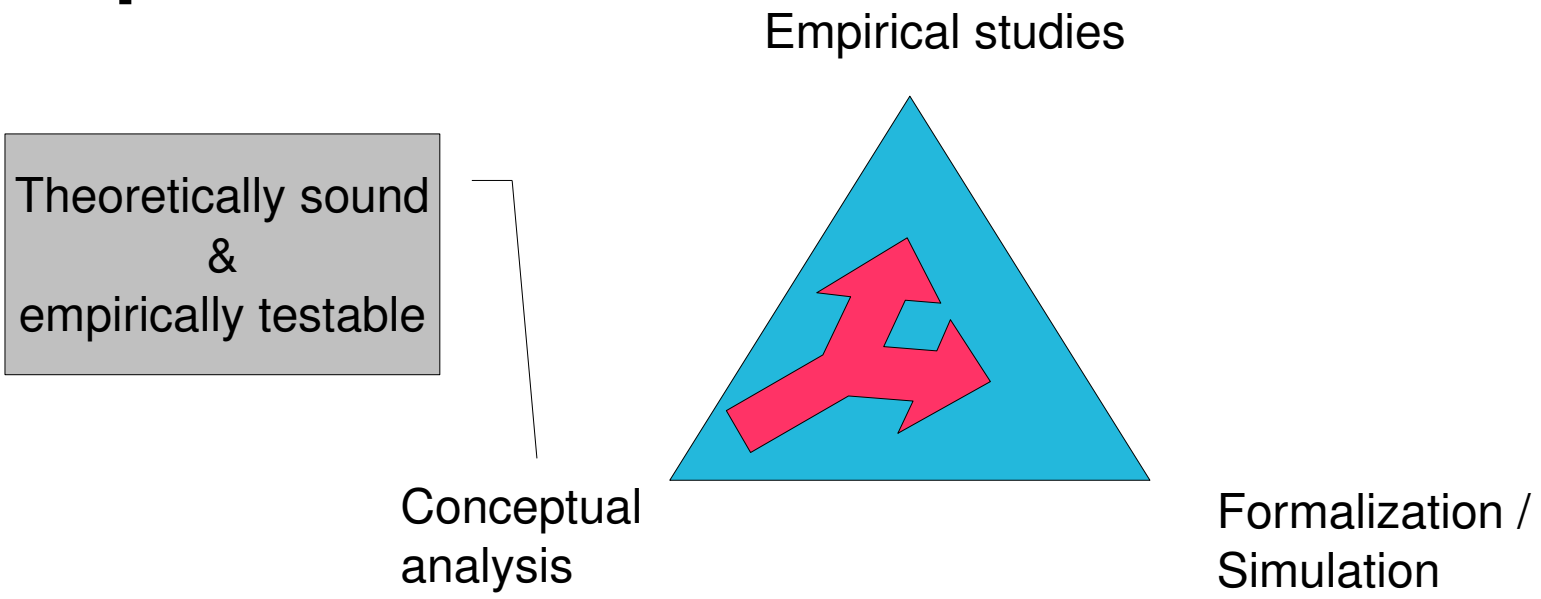
Output:



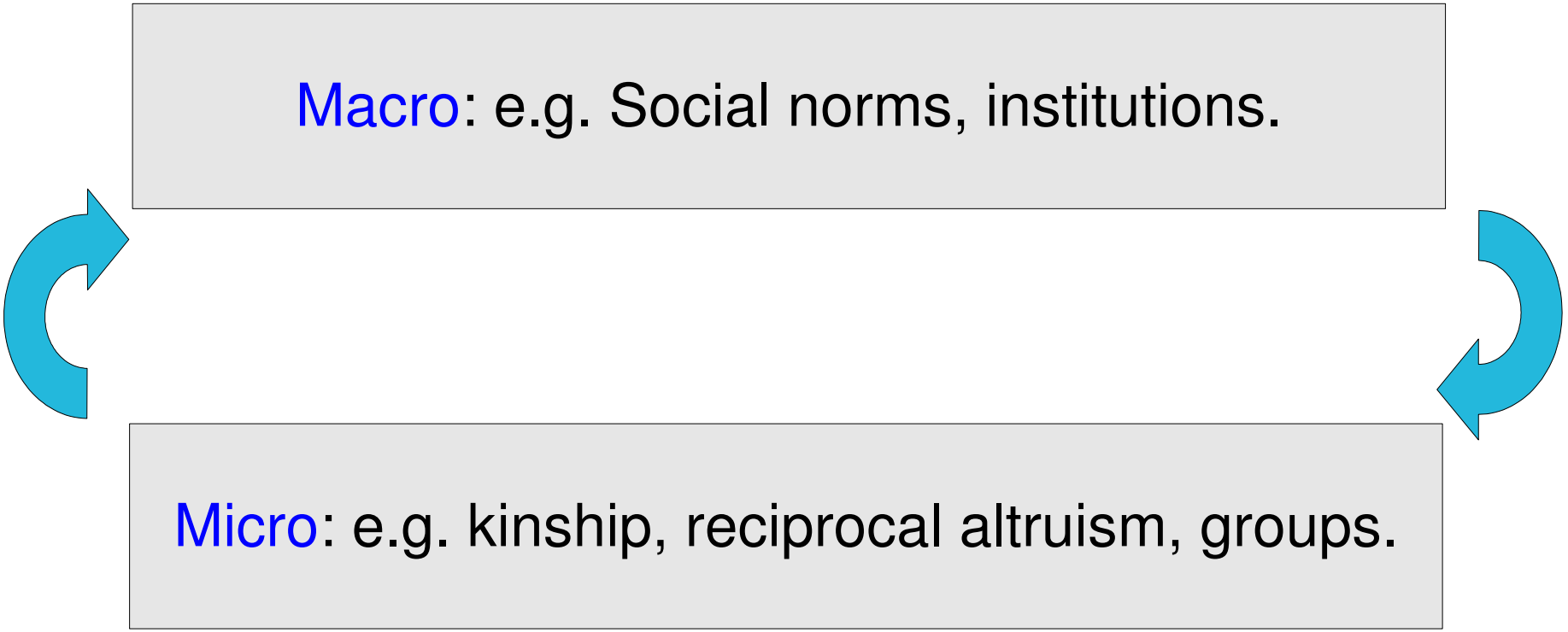
Evolution of cooperation

Input: Contribution from different fields of research.

Output:

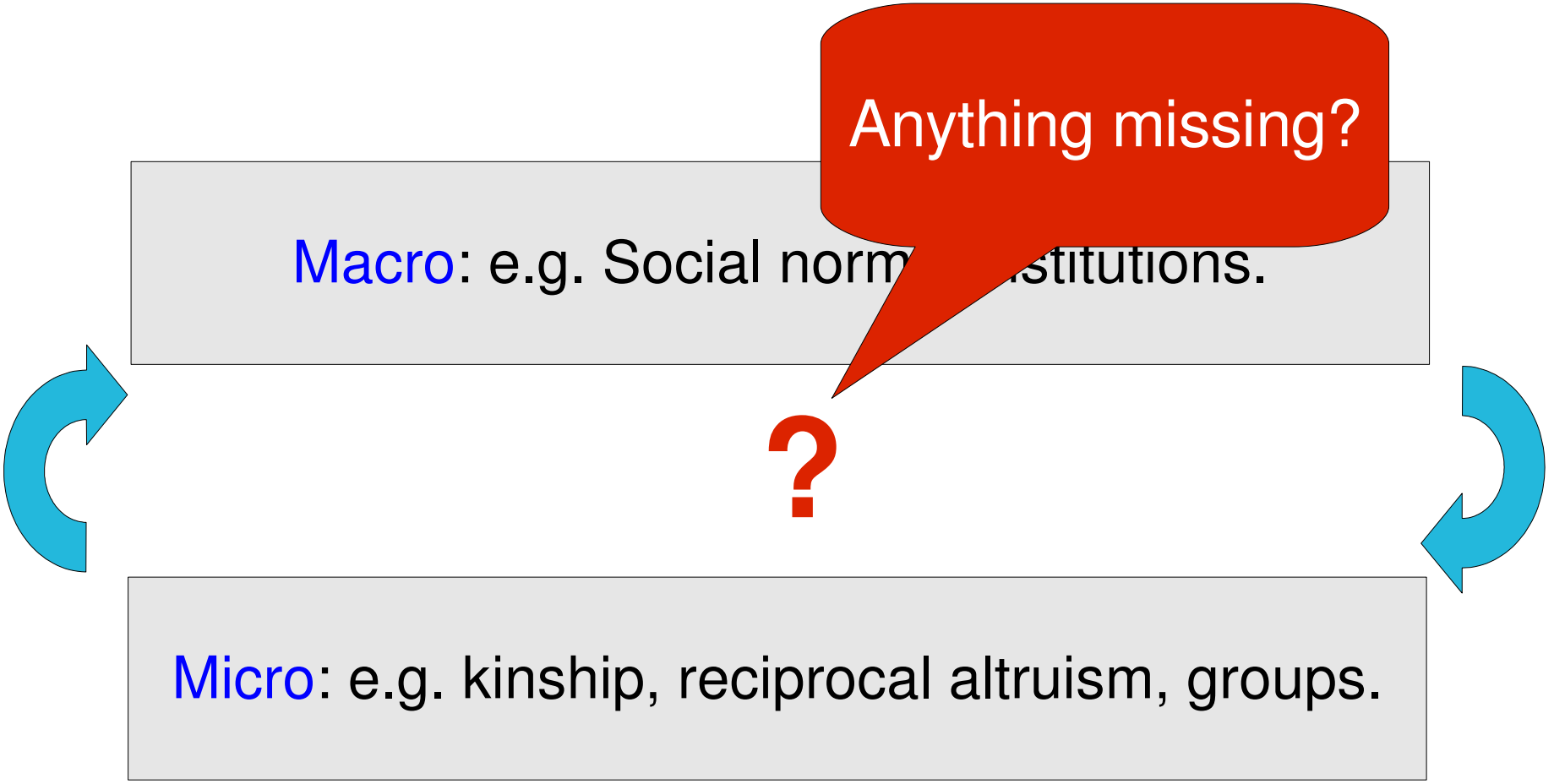


Traditional explanations



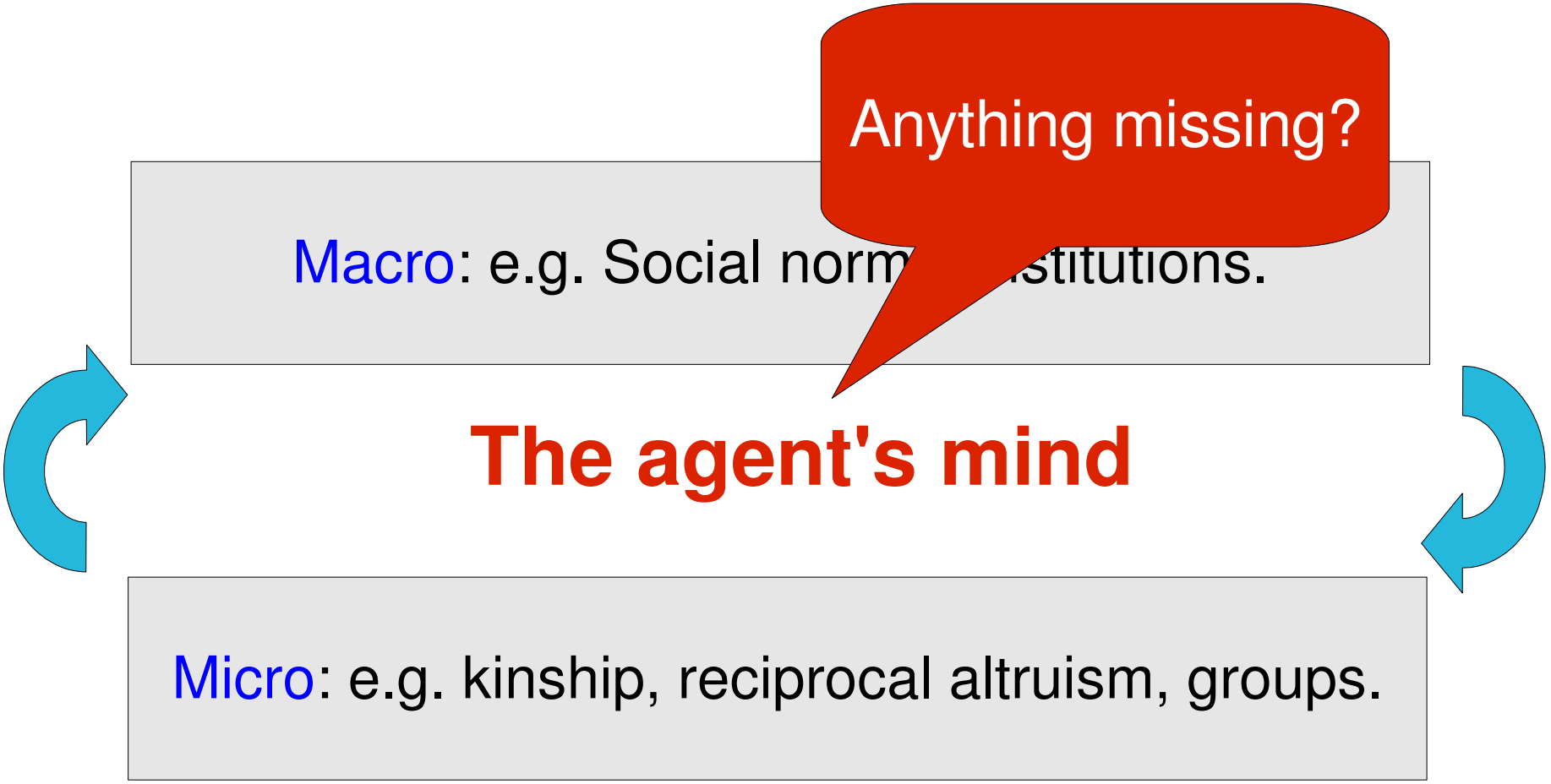
Problem: cooperation in large, non-kin groups

Traditional explanations



Problem: cooperation in large, non-kin groups

Traditional explanations



Problem: cooperation in large, non-kin groups

A possible (modular) view of the mind

- Accredited tradition in evolutionary psychology describes the mind as an “adaptive toolbox” (Cosmides, Tooby, Barkow, The Adapted Mind, 1992).
- Each tool in the box is a context-specific module, evolved to solve specific problems of adaptation (e.g. partner selection, cheater detection in trading, etc).
- *Alternative view: explain behavior as the effect of general, high-level cognitive faculties, i.e. processing of mental representations, beliefs transmission, learning. (Is it evolutionary plausible?)*

Autonomous cognitive agents

- Beliefs and Goals: mental representations with a regulatory function over the behavior of cognitive agents.
- Goal generation rule: every new goal is a mean for an existing goal of the agent. (**e. g.** *Agent X has the goal P; X believes that Q is a mean for P; then X has the goal Q*.)
- Goal's adoption: an agent is not equipped with the means to fulfill all is possible goals. Sociality offers the opportunity to increase the agents' power, by having their goals been adopted by another agent.

The mechanism of goal's adoption

- Agent X has an adopted goal G:
  1. If X believes that Y has such goal G;
  2. As long as X believes that Y has the goal G (*adoption is not imitation*);
  3. In order to have G satisfied for Agent Y.
- Taking Y's point of view: if Y is able to force X to pay the cost required by G, he can exploit Y's power. Otherwise Y must provide X with the necessary motivation for G (e.g. exchange, trade)

Types of goal's adoption

- X can adopt Y's goal G:
  - **Instrumentally**: in order to achieve another goal G' (G is mean for another of X's goal);
  - **Cooperatively**: when G' is a common goal, shared by both agent X and Y;
  - **Altruistically**: when X's adoption is not motivated by any other goals of X.
- These categories are relevant only if look at the agent's mind. If agent's actions affect biological or social function (goals outside the mind of the agent), we can calculate actions' reward in terms of inclusive fitness or reciprocal altruism (biological function), or as group membership (social function).

Is altruism a paradox?

- The fact that an agent is self-interested just means that he is goal-autonomous, that is, he pursues his own, internally represented purposes.
- To be altruist does not mean not to be an autonomous decision-maker; and does not mean to be irrational (hence the paradox of altruistic behavior).
- Being self-interested is not the same of being selfish (the utilitarian approach mixes up the two notions).

Is altruism a paradox? (2)

- Psychology usually accept a weaker notion of altruism, a pseudo- or limited altruism. It assumes that every time I made an action which is beneficial for another agent, I can and will always obtain a positive reward (internal or external).
- This notion is consistent with classical game-theoretic models, where agents are utility-maximizers.
- At the cognitive level is necessary to keep separated (positive) unexpected rewards from intended ones, expectations; and distinguish between expectations and motivations.

Reciprocity

- Using the notion of goal adoption we can define an action as reciprocity among two agents:
  - X adopts Y's goal P and
  - Y adopts X's goal Q **because of** the previous adoption from X.
- Y's adoption is instrumental to another goal of agent Y: to reduce the perceived unbalance between his present state and the state of agent X.

For a gradualistic approach to pro-social behavior (1)

- Example:
  - The CIA gives money to Al Qaeda, because US has great interests in the middle-east region.
  - Al Qaeda uses US dollars to pursues his own ends, which – indirectly – advantage US politics.
- It seems inappropriate to describe the example above as a cooperative social interaction.
- Using the tools of classical game theory we have to analyze in the same way the above example as well as a real case of reciprocal altruism.
- Thinking in terms of rewards, everything is reciprocal altruism.

For a gradualistic approach to pro-social behavior (2)

- At the behavioral level we cannot distinguish reciprocity from a mere donation, cheating from punishment, a social norm from a behavioral equilibrium.
- Concepts like reciprocity, cooperation and altruism should be defined at the mental level.
- What we gain? A more sophisticated teleonomic model can be used to say:
  - When the simple mechanism of reproduction with mutation is sufficient to explain the phenomena.
  - where - at which level - the selection takes place.
  - how higher levels of organization emerge from lower ones (e.g. the social from the individual level, the institutional from the social, ...)

Advantages of Present Study

- Simulation are used to check consistency of conceptual analysis.
- Model is anchored to real data (taken from biological literature).
  - Reduce arbitrariness of multi-parametered simulations.
- Main theoretical questions addressed:
  - Can a non-instrumental forms of altruism survive in a system that assign positive rewards to utility-maximizers?

The Problem in Nature

- Vampire bats' altruism (Wilkinson, 1984):
  - If successful, hunters regurgitate in favour of unlucky fellows
  - Survival rate
    - rises up to 80% of the initial population,
    - as opposed to 20% without altruism.
- Why?
  - Kin selection.
    - Not quite: low average rate of relatedness among individuals living in the same roosts (around 6%).
  - Reciprocal altruism.

The simulation model (1)

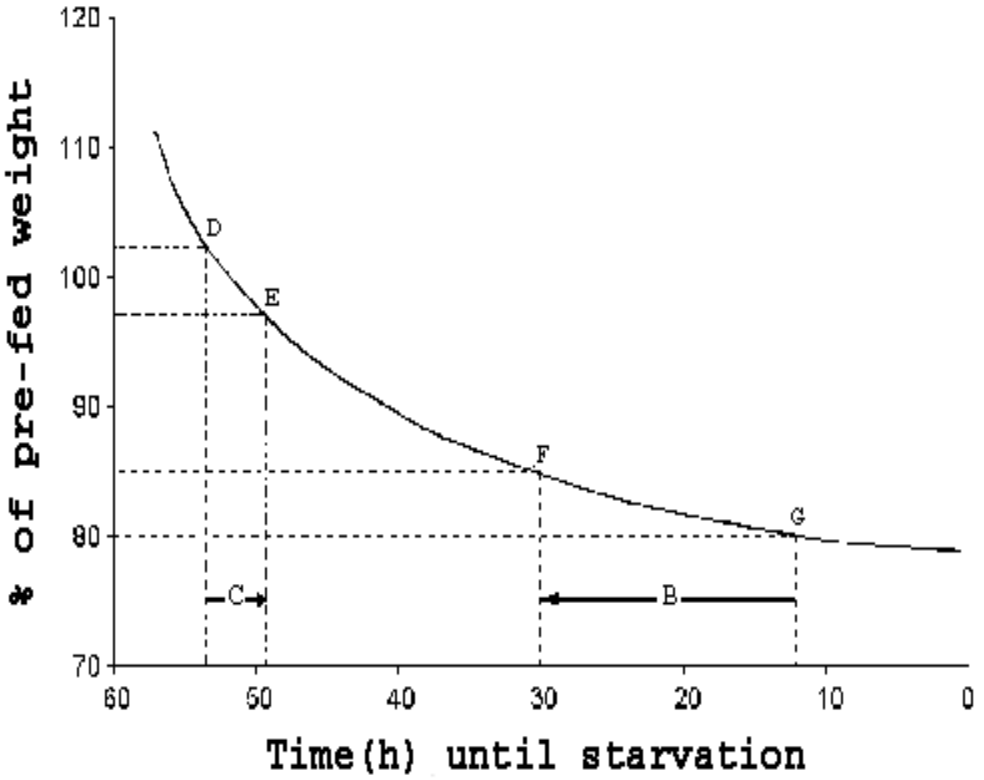
- Parameters (as in nature):
  - Roosts: a. 15 individuals each.
  - Food:
    - 93% individuals hunt each night.
    - 7% starve and ask for help.

- Agents have goals:
    - survival and
    - normative (give help).

  with variable strength.
  - Five patterns of goal relationships emerge:
    - NG is null: *cheaters.*
    - SG is null: *martyrs.*
    - Gs are equal: *fair.*
    - NG is stronger: *generous.*
    - SG is stronger: *prudent* (simple-like).

The simulation model (2)

- Night: hunting (if good, 60hs autonomy).
- Day:
  - Grooming: increase familiarity and reciprocity -> Credit network: initialised with food-sharing
    - Donation = credit.
    - Reciprocity extinguishes credits.
    - Investigated when asked for help: no more than two consecutive donations per postulant are allowed.
  - (possibly) food-sharing: spare life of starving bats.
  - Reproduction (cloning every 2 mature individuals).
  - If following two unsuccessful nights and not helped, Death after 12 hs starvation.

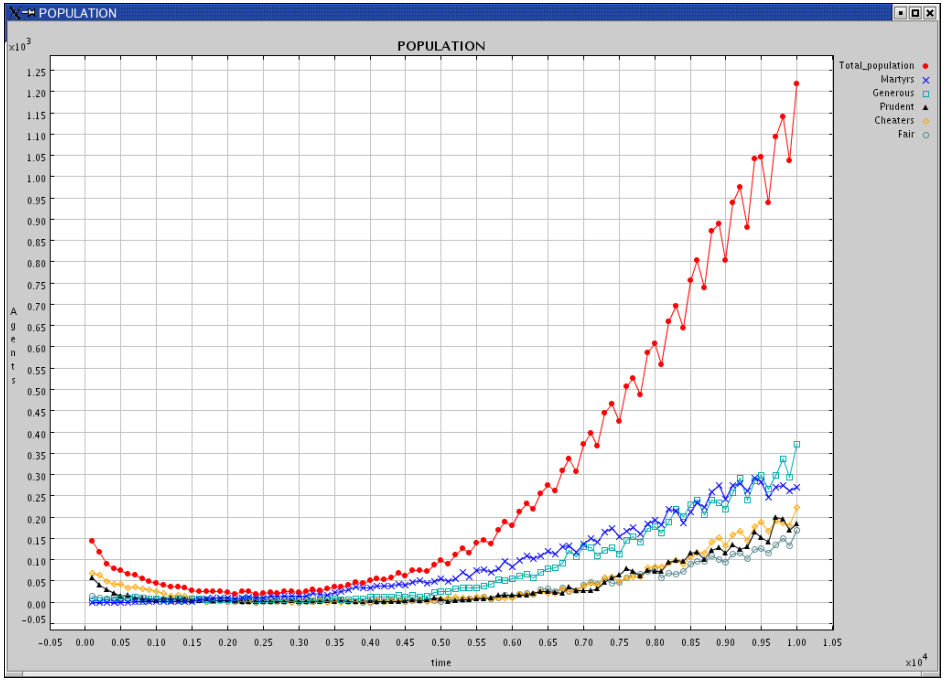


Dynamic vampire bats

- Goal-dynamics: an essential aspect of cognition (Conte, 2000):
  - Based on beliefs, goals are generated, interrupted, dropped etc. as an effect of value change.
- How does goal value change?
  - In our simulations, NG
    - Grows with donations received
    - Decreases with unreciprocated donations
- Essentially, keeping constant for simplicity SG, the strategy played by an agent is caused by the dynamics of its NG.

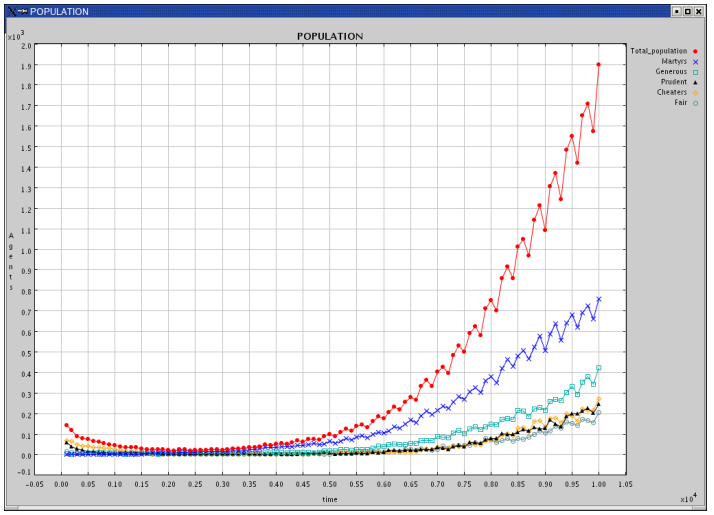
Findings (1): strategies emerge

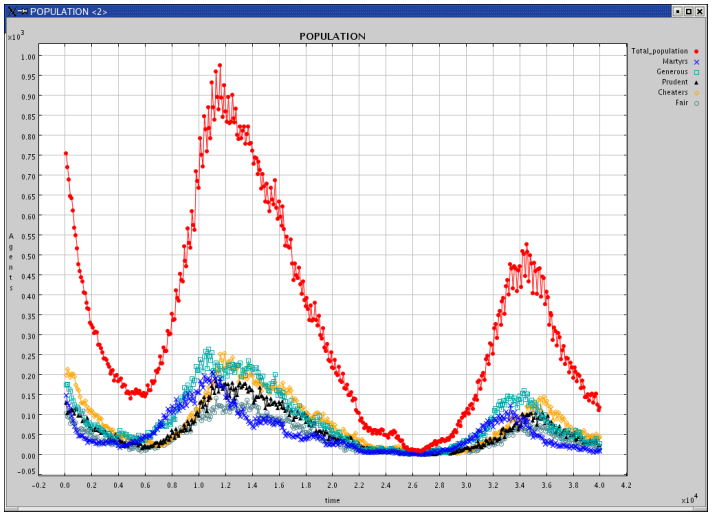
- With 10 roosts of 15 dynamic smart agents each (same conditions as in previous experiment), starting prudent and cheaters, 40 ys (4 gen.)
- Strategies differentiate
  - Altruistic ones being fitter than cheating,
  - Cheating is kept inoffensive (never grows even when cheaters are a majority)
  - Difference increases with inheritance of goal-value.



Findings (2). dominance of martyrdom

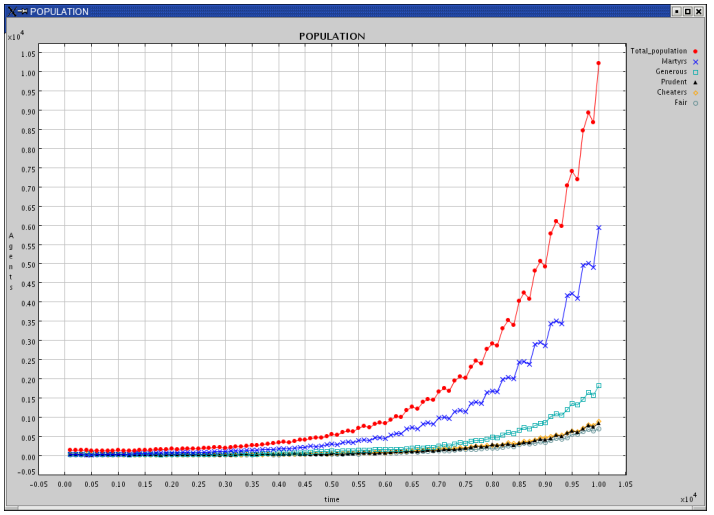
- Martyrdom is dominant, because it is self-reinforcing:
  - if donations
    - **increase**, NG grows *ad libitum.*
    - **decrease**, NG cannot lower below a given threshold.
  - with an upper limit,
    - it behaves like other straegies.
    - population varies periodically.

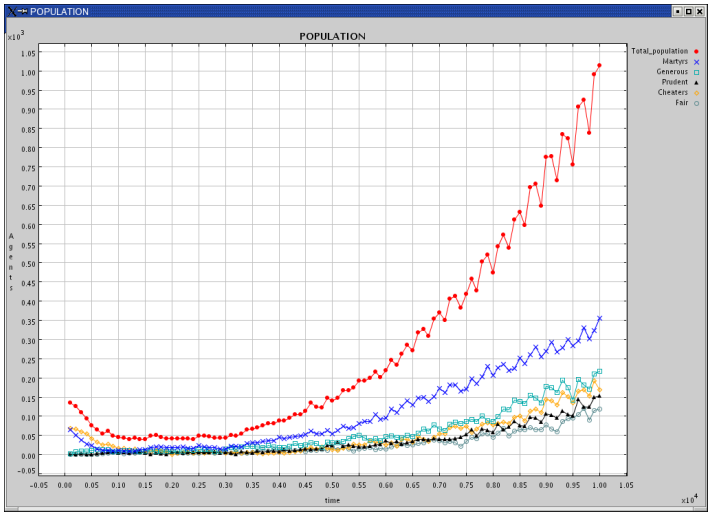




Findings (3). initial strategy does not matter

- Findings are stable while initialising simulations with different strategies
- The most altruistic strategies are always dominant





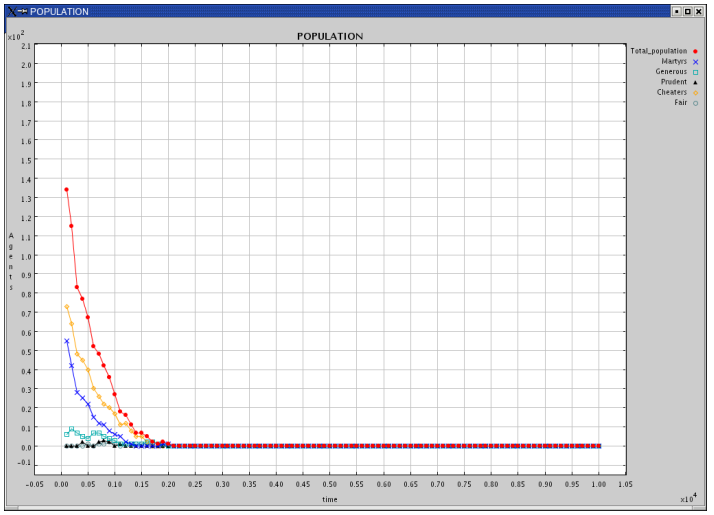
Graph

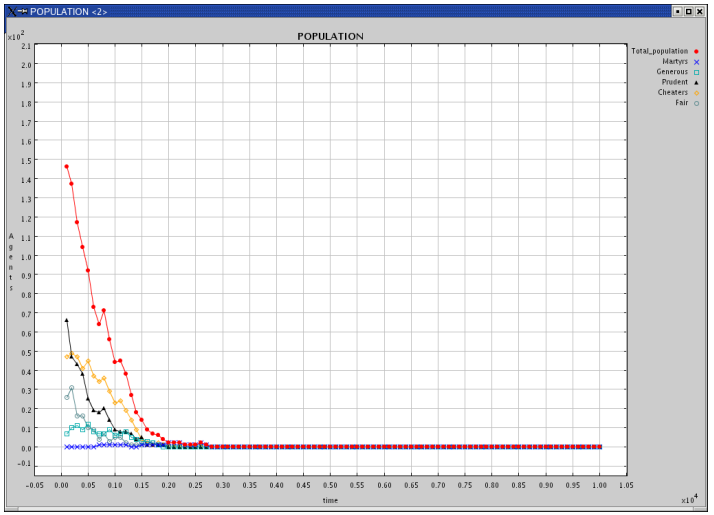
Above: all strategies present at the beginning of simulation

Below: martyrs & cheaters at start

Findings (4). retaliation severity matters

- With more severe retaliation (debtors are denied help until credits are extinguished)
- Population extinguishes





Graph

Above: all strategies present at the beginning of simulation

Below: prudents & cheaters at start

To sum up

- Artificial vampire bats are only a partial implementation of the theory of social action.
- The theory should be used to generate new experimental hypothesis (for artificial, as well as natural subjects).