In: Proceedings of the 3rd European Conference on Cognitive Science, ECCS99, Sienna, Italy, 1999, pages 57-62.

# On being responsible: how to be individualistic and smile benevolence to the others

**Nuno David**
ISCTE/DCTI
Department of Information Sciences and Technologies
Av. das Forças Armadas
1600 Lisbon, Portugal
+351 1 7903911
Nuno.David@iscte.pt
http://www.iscte.pt/~nmcd

**Jaime Simão Sichman**
Dept. of Computer Engineering
University of São Paulo
Av. Prof. Luciano Gualberto, 158, tv. 3
05508-900 São Paulo, Brazil
+55 11 8185397
jaime@pcs.usp.br
http://www.pcs.usp.br/~jaime

**Helder Coelho**
Dept. of Computer Science
University of Lisbon
Bloco C5, Piso 1
Campo Grande
1700 Lisbon, Portugal
+351 1 7500122
hcoelho@di.fc.ul.pt
http://www.di.fc.ul.pt/areas/ai

**ABSTRACT**
In this paper, we discuss the balancing problem relative to the agents' individual and global social achievements in Multi-Agent Systems. Our claim is that any social principle prescribed within a level of abstraction above Newell's Knowledge Level, in order to balance the needs of the individual and the system as a whole, can not adopt an utility-oriented scale of performance but instead must adopt a *substantialist*-oriented (goal-driven) concept of performance. On this perspective, we propose a number of goal-generating principles for social responsible agents, towards an uncompromising individualistic view of autonomous agents that nevertheless attempt to motivate the welfare of collectives.

**Keywords**
Multi-Agent Systems, benevolence/non-benevolence, individual and social performance, social rationality

## BENEVOLENCE DILEMMAS, *DO NOT SMILE*

In *Multi-Agent Systems* (MAS) agents are desirably autonomous, although invariably facing the matter of fact of limited self-sufficiency. Autonomy expresses the agents' ability to act without direct intervention or guidance of others (e.g. humans), primarily according to their adopted goals, relevant competence, and specific problem solving abilities. In addition, while not compromising their autonomy, inasmuch as agents may use each others' capabilities and resources to achieve their ends, agent are also and necessarily interdependent [Castelfranchi, 1990].

Notwithstanding, autonomy and lack of self-sufficiency may even entitle agents, deliberately and selectively, not to interact with others (e.g. co-operate) exclusively or inasmuch as they do not expect to benefit from such pro-activity or external request. In this sense, the social interaction model is a *bottom-up* one: social interactions and organisations are produced as a result of the agents' efforts to achieve their own goals [Conte and Sichman, 1995]. Interactions like co-operation are not assumed *a priori* but the principle of *non-benevolence* is presumed: agents are not prescribed to help each other but decide autonomously whether or not to co-operate with others. As pointed out by Jennings and Campos [1997], in this *contructist* view of multiple agent systems, the measure of *performance* of the system is fundamentally focused on the *individual performance* of the agent, nearly disregarding the achievements of other agents or the overall agency

In effect, if the principle of *non-benevolence* is assumed, we may find two major trends for measuring the agents' *individual performance*.

The first one adopts a utility oriented scale, calculated according to the cost/benefit of the agents' actions against the worth of the corresponding goals, whatever goals these may be. The agents' decision functions are fundamentally concerned with choosing actions that maximise individual utility, according to the classic principle of economic rationality ([e.g. [Zotkin and Rosenchein, 1994; Sandholm and Lesser, 1995]).

On the other hand, an inclined socio-psychological perspective of *performance* will most probably value a *substantialist* [Conte and Pedone, 1998] view of rationality; that is, individual performance measured in terms of the agents' attained goals. Here, similar to Newell's principle of rationality [Newell, 1982], the real *motive* for being rational is focused on the agents' own goals. In this case, the agents' decision functions are essentially concerned with the choice of adequate partners (organisations) and relevant actions in order to better achieve their goals, according to different patterns of dependence and influencing power over the others [Castelfranchi *et al.*, 1992].

On the other extreme of the benevolence vector, the earlier *reductionist* view [Jennings and Campos, 1997]