Valorizing Prejudice in MAS: A Computational Model

Rino Falcone
ISTC-CNR
Rome, Italy
rino.falcone@istc.cnr.it

Alessandro Sapienza
ISTC-CNR
Rome, Italy
alessandro.sapienza@istc.cnr.it

Cristiano Castelfranchi
ISTC-CNR
Rome, Italy
cristiano.castelfranchi@istc.cnr.it

ABSTRACT

In MAS studies on Trust building and dynamics the role of direct/personal experience and of recommendations and reputation is proportionally overrated; while the importance of inferential processes in deriving the evaluation of trustees' trustworthiness is underestimated and not exploited.

In this paper we focus on the importance of generalized knowledge: agents' categories. The cognitive advantage of generalized knowledge can be synthesized in this claim: "It allows us to know a lot about something/somebody we do not directly know". At a social level this means that I can know *a lot of things on people that I never met*; it is social "prejudice" with its good side and fundamental contribution to social exchange. In this study we experimentally inquire the role played by categories' reputation with respect to the reputation and opinion on single agents: when it is better to rely on the first ones and when are more reliable the second ones. Our claim is that: *the larger the population and the ignorance about the trustworthiness of each individual* (as it happens in an open world) *the more precious the role of trust in categories*.

This powerful inferential device has to be strongly present in WEB societies supported by MAS.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**] : Distributed Artificial Intelligence - *multiagent systems*

General Terms

Experimentation, Human Factors, Reliability, Theory

Keywords

Trust and reputation, Cognitive models, Social simulation

1. INTRODUCTION

In MultiAgent Systems (MAS) and Online Social Networks (OSN) studies on Trust building and dynamics the role of direct/personal experience and of recommendations and reputation (although important) is proportionally overrated; while the importance of inferential processes in deriving the evaluation of trustee's trustworthiness is underestimated and not sufficiently exploited (a part from the so called "transitivity", which is also, very often, wrongly founded).

In particular, generalization and instantiation from classes and categories [8], and analogical reasoning (from task to task and from agent to agent) really should receive much more attention. In this paper we focus on the importance of generalized knowledge: agents' categories. The cognitive advantage of generalized knowledge (building *classes*, *prototypes*, *categories*, etc.), can be synthesized in this obvious claim: "It allows us to know a lot about something/somebody we do not directly know" (for example, I never saw Mary's dog, but - since it is *a dog* - I know hundreds of things about it).

At a social level this means that I can know *a lot of things on people that I never met*; it is social "prejudice" with its good side and fundamental contribution to social exchange. How can I trust (for drugs prescription) a medical doctor that I never met before and nobody of my friends knows? Because he is a doctor!
Of course we are underlining the positive aspects of generalized knowledge, its essential role for having information on people never met before and about whom no one gave testimony. The more rich and accurate this knowledge is, the more it is useful. It offers huge opportunity both for realizing productive cooperation and for avoiding risky interactions. The problem is when the *uncertainty about the features* of the categories is too large or it is too wide the *variability of the performers* within them. In our culture we attribute a negative sense to the concept of prejudice, and this because we want underline how generalized knowledge can produce unjust judgments against individuals (or groups) when superficially applied (or worst, on the basis of precise discriminatory intents). Here we want rather point out the positive aspects of the prejudice concept.

In this study we intend to explain and experimentally show the advantage of trust evaluation based on classes' reputation with respect to the reputation and opinion on single potential trustees (partners). In an open world or in a broad population how can we have sufficient direct or reported experience on everybody? The quantity of potential trustees in that population or net that might be excellent partners but that nobody knows enough can be high.

Our claim is that: *the larger the population and the ignorance about the trustworthiness of each individual the more precious the role of trust in categories*. If I know (through signals, marks, declaration, ...) the class of a given guy/agent I can have a reliable opinion of its trustworthiness derived from its class-membership.
It is clear that the advantages of such cognitive power provided by categories and prejudices does not only depend on recommendation and reputation about categories. We can personally build - by generalization - our evaluation of a given category from our direct experience with its members (this is fact happens in our experiments for the agents that later have to propagate their recommendation about). However, in this simulation we have in the trustor (which has to decide whom rely on) only a prejudice based on recommendations about that category and not its personal experience.

After a certain degree on direct experiences and circulation of recommendations, the performance of the evaluation based on classes will perform better; and in certain cases there will be no alternative at all: we do not have any evaluation on that individual, a part from its category; either we work on inferential instantiation of trustworthiness or we loose a lot of potential partners. This powerful inferential device has to be strongly present in WEB societies supported by MAS. We simplify here the problem of the generalization process, of how to form judgement about groups, classes, etc. by putting aside for example inference from other classes (higher or sub); we build opinion (and then its transmission) about classes on the bases of experience with a number of subjects of a given class.
First of all, we want to clarify that here we are not interested in steretypes, but in categories. We define steretypes as the set of features that, in a given culture/opinion, characterize and distinguish that specific group of people.

Knowing the stereotype of an agent could be expensive and time consuming. Here we are just interested in the fact that an agent belongs to a category: it has not to be a costly process and the recognition must be well discriminative and not-cheating. There should be visible and reliable "signals" of that membership. In fact, the usefulness of categories, groups, roles, etc. makes fundamental the role of the *signs* for recognizing or inferring the category of a given agent. That's why in social life are so important coats, uniforms, titles, badges, diplomas, etc. and it is crucial their exhibition and the assurance of their authenticity (and, on the other side, the ability to falsify and deceive). In this preliminary model and simulation let us put aside this crucial issue of indirect competence and reliability *signaling*; let us assume that the membership to a given class or category is true and transparent: the category of a given agent is public, common knowledge.

Differently from [2][11][18], in this work we do not address the problem of learning categorical knowledge and we assum that the categorizzation process is objective.

Similarly to [3], we give agents the possibility to recommend categories and this is the key point of this paper.

In the majority of the cases available in the literature, the concept of recommendation is used concerning recommender systems [1]. These ones can be realized using both past experience (content-based RS) [14] or collaborative filtering, in which the contribute of single agents/users is used to provide group recommendations to other agents/users.

Focusing on collaborative filtering, the concepts of similarity and trust are often exploited (together or separately) to determine which contributes are more important in the aggregation phase [15][19]. For instance, in [7] authors provide a system able to recommend to users group that they could join in Online Social Network. Here it is introduced the concepts of *compactness* of a social group, defined as the weighted mean of the two dimensions of similarity and trust.

Even in [12] authors present a clustering-based recommender system that exploits both similarity and trust, generating two different cluster views and combining them to obtain better results.

Another example is [6] where authors use information regarding social friendships in order to provide users with more accurate suggestions and rankings on items of their interest.

A classical decentralized approach is referral systems [21], where agents adaptively give referrals to one another.

Information sources come into play in FIRE [13], a trust and reputation model that use them to produce a comprehensive assessment of an agent's likely performance. Here authors take into account open MAS, where agents continuously enter and leave the system. Specifically, FIRE exploits interaction trust, role-based trust, witness reputation, and certified reputation to provide trust metrics.

The described solutions are quite similar to our work, although we contextualized this problem to information sources. However we do not investigate recommendations with just the aim of suggesting a particular trustee, but also for inquiring categories' recommendations.

2. RECOMMENDATION AND REPUTATION: DEFINITIONS

Let us consider a set of agents $Ag_1, ..., Ag_n$ in a given world (for example a social network). We consider that each agent in this world could have trust relationships with anyone else. On the basis of these interactions the agents can evaluate the trust degree of their partners, so building their judgments about the trustworthiness of the agents with whom they interacted in the past.

The possibility to access to these judgements, through recommendations, is one of the main sources for trusting agents outside the circle of closer friends. Exactly for this reason recommendation and reputation are the more studied and diffused tools in the trust domain [16].

We define

$$\mathrm{Rec}_{x,y,z}(\tau) \qquad (1)$$

where $x, y, z \in \{Ag_1, Ag_2, ..., Ag_n\}$, we call D the specific domain: $D \equiv \{Ag_1, Ag_2, ..., Ag_n\}$

and $0 \le \mathrm{Rec}_{x,y,z}(\tau) \le 1$

τ, as established in the trust model of [4], is the task on which the recommender expresses the evaluation about y.

In words: $\mathrm{Rec}_{x,y,z}(\tau)$ is the value of x's recommendation about y performing the task τ, where z is the agent receiving this recommendation. In this paper, for sake of simplicity, we do not introduce any correlation/influence between the value of the recommendations and the kind of the agent receiving it: the value

of the recommendation does not depend from the agent to whom it is communicated.

So (1) represents the basic expression for recommendation.

We can also define a more complex expression of recommendation, a sort of *average recommendation*:

$$\sum_{x=Ag_1}^{Ag_n} \mathrm{Re}\, c_{x,y,z}(\tau) / n \qquad (2)$$

in which all the agents in the domain express their individual recommendation on the agent *y* with respect the task τ and the total value is divided by the number of agents.

We consider the expression (2) as the reputation of the agent y with respect to the task τ in the domain D.

Of course the reputation concept is more complex than the simplified version here introduced [5][17].

It is in fact the value that would emerge in the case in which we receive from each agent in the world its recommendation about y (considering each agent as equally reliable).

In the case in which an agent has to be recommended not only on one task but on a set of tasks (τ_1 , ..., τ_k), we could define instead of (1) and (2) the following expressions:

$$\sum_{i=1}^{k} \mathrm{Re}\, c_{x,y,z}(\tau_i) / k \qquad (3)$$

that represents the *x*'s recommendation about *y* performing the set of tasks (τ_1,...,τ_k), where *z* is the agent receiving this recommendation.

Imagine having to assign a meta-task (composed of a set of task) to one of several agents. In this case the information given from the formula (3) could be useful for selecting on average (with respect to the tasks) the more performative one.

$$\sum_{x=Ag_1}^{Agn} \sum_{i=1}^{k} \mathrm{Re}\, c_{x,y,z}(\tau_i) / nk \qquad (4)$$

that represents a sort *of average recommendation* from the set of agents in D, about *y* performing the set of tasks (τ_1 , ..., τ_k). *We consider the expression (4) as the reputation of the agent y with respect the set of tasks (τ_1 , ...,τ_k), in the domain D.*

Having to assign the meta-task proposed above, the information given from the formula (4) could be useful for selecting on average (with respect to both the tasks and the agents) the more performative one.

2.1 Using Categories

As described above, an interesting approach for evaluating agents is to classify them in specific categories already pre-judged/rated and as a consequence to do inherit to the agents the properties of their own categories.

So we can introduce also the *recommendations about categories*, not just about agents (we discuss elsewhere how these recommendations are formed). In this sense we define:

$$\mathrm{Re}\, c_{x,Cy,z}(\tau) \qquad (5)$$

where $x \in \{Ag_1, Ag_2, ..., Ag_n\}$ and

$C_y \subseteq \{Ag_1, Ag_2, ..., Ag_n\}$,

$0 \le \mathrm{Re}\, c_{x,Cy,z}(\tau) \le 1$

In words: $\mathrm{Re}\, c_{x,Cy,z}(\tau)$ *is the value of x's recommendation about the agents included in category C_y when they perform the task τ,* (as usual *z* is the agent receiving this recommendation).

We again define a more complex expression of recommendation, a sort of *average recommendation*:

$$\sum_{x=Ag_1}^{Ag_n} \mathrm{Re}\, c_{x,Cy,z}(\tau) / n \qquad (6)$$

in which all the agents in the domain express their individual recommendation on the category C_y with respect the task τ and the total value is divided by the number of agents.

We consider the expression (6) as the reputation of the category C_y with respect the task τ in the domain D.

Now we extend to the categories, in particular to C_y, the recommendations on a set of tasks (τ_1, ..., τ_k):

$$\sum_{i=1}^{k} \mathrm{Re}\, c_{x,Cy,z}(\tau_i) / k \qquad (7)$$

that represents *the value of x's recommendation about the agents included in category C_y when they perform the set of tasks (τ_1,...,τ_k).*

Finally, we define:

$$\sum_{x=Ag_1}^{Agn} \sum_{i=1}^{k} \mathrm{Re}\, c_{x,Cy,z}(\tau_i) / nk \qquad (8)$$

that represents *the value of the reputation of the category C_y (of all the agents y included in C_y) with respect the set of tasks (τ_1,...,τ_k), in the domain D.*

2.2 Definitions of Interest for this Work

In this paper we are in particular interested in the case in which *z* (a new agent introduced in the world) asks for recommendation to *x* ($x \in D$) about an agent belonging to its domain *D* (the set of all the agents in the world) for performing the task τ. *x* will select the best evaluated y, with $y \in D_x$ on the basis of formula:

$$\max_{y \in D_x} (\mathrm{Re}\, c_{x,y,z}(\tau)) \qquad (9)$$

where $D_x \equiv \{Ag_1, Ag_2, ..., Ag_m\}$, D_x includes all the agents evaluated by *x*. They are a subset of *D*: $D_x \subseteq D$.

In general *D* and D_x are different because *x* does not necessarily know (has interacted with) all the agents in *D*.

z asks for recommendations not only to one agent, but to a set of different agents: $x \in D_z$, and selects the best one on the basis of the value given from the formula:

$$\max_{x \in D_z}(\max_{y \in D_x}(\mathrm{Re}\, c_{x,y,z}(\tau)))$$

(10)

$D_z \subseteq D$, z could ask to all the agents in the world or to a defined subset of it (see later).

We are also interested to the case in which *z* ask for recommendations to *x* about a specific *agents' category* for performing the task τ. *x* has to select the best evaluated C_y among the different C_y x has interacted with (we are supposing that each agent in the world *D*, belongs to a category C_y in the set $\{C_{y1}, C_{y2}, \ldots, C_{yn}\}$).

In this case we have the following formulas:

$$\max_{Cy \in D_x}(\mathrm{Re}\, c_{x,Cy,z}(\tau)) \quad (11)$$

that returns the category best evaluated from the point of view of an agent (x). And

$$\max_{x \in D_z}(\max_{Cy \in D_x}(\mathrm{Re}\, c_{x,Cy,z}(\tau))) \quad (12)$$

that returns the category best evaluated from the point of view of all the agents included in D_z.

3. COMPUTATIONAL MODEL

3.1 NetLogo

In order to realize our simulations, we exploited the software NetLogo [20]. It is an open source agent-based programming environment written in Java, particularly suited for modeling natural and social phenomena.

In NetLogo everything is an agent (also the patches that compose the world in which the other agents move) and it is possible to create and model many kind of them, specifying how they relate to each other and giving individual instructions. It is also possible to modify the world at run time, to further answer those "what if" questions that pop up while investigating the models.

It splits the programming part, in which the programmer can set up the environment of the simulation and specify the behavior of turtles, and the visual part, in which the user can start the simulation, control it changing its parameters and see the result at run time, through the view representing the world, plots and output monitors.

Although NetLogo is an excellent instrument for simulation's tasks, it is devoid of adequate computational libraries to implement the computational model of trust on information's sources. Then it has proved necessary to expand it with a Java plug-in made by us, able to fill these gaps. In practice, this trust plug-in implements all the model of trust on information's sources.

3.2 General Setup

In every scenario there are four general categories, called A,B,C and D, each one characterized by:

1. an **average value of trustworthiness**, in range [0,100];
2. an **uncertainty value**, in range [0,100].

Those two values are exploited to generate the **objective trustworthiness** of each trustee, defined as *the probability that, concerning a specific kind of required information, the trustee will communicate the right information*.

Of course, the trustworthiness of categories and trustees is strongly related to the kind of requested information/task. In these simulations we use just one kind of information in which the categories A, B, C and D have 80, 60, 40 and 20% of average value of trustworthiness respectively. The uncertainty value is fixed to 20% for all of them.

The simulations were carried out using two different numbers of trustee: 20 trustees for each category and 100 trustees for each category. In both cases we used just one trustor.

3.3 How the simulations work

Simulations are mainly composed by two main steps that repeat continuously. In the first step, called **exploration phase**, agents move into the world asking to their neighbors (other agents with a distance of less than 3 NetLogo patches) for the information P. Then they memorize the performance of each neighbor both as individual element and as a member of its own category.

The performance of a agent can assume just the two values 1 or 0, with 1 meaning that the agent is supporting the information P and 0 meaning that it is opposing to P. For sake of simplicity, we assume that P is always true.

We also choose to let agents move with a probability of 10% (each agent moves, with a probability of 10%, one patch in a random direction) so, on the one hand we can say that the agents change their neighbors after each tick, but, on the other hand this change is quite slow and, given the number of ticks realized they are not able to know all the other agents in the world, but they know properly just a subset of them.

We call the set of neighbors with whom agents interact in each tick: their *neighborhood*.

The exploration phase has a variable duration, going from 100 ticks to 1 tick. Depending on this value, agents will have a better or worse knowledge of their neighborhoods.

Then, in a second step (**querying phase**) we introduce in the world a trustor (a new agent with no knowlegde about the trustworthiness of other agents and categories, and that has the necessity to trust someone reliable for a given task). It will select a given subset of the population and it will query them. In particular, the trustor will ask them for the best category and the best trustee they have experienced.

In this way, the trustor is able to collect information about the best recommended category and agent.

It is important to underline that the trustor is collecting information from the agents considering them as equally trustworthy with respect to the task of "providing recommendations". Otherwise it should weigh differently these recommendations.

Then it will select the nearest agent belonging to the best recommended category and it will compare it, in terms of objective trustworthiness, with the best recommended individual agent (trustee).

The possible responses are:

- **trustee wins**: the trustee selected with individual recommendation is better than the one selected by the means of category; then this method gets one point;
- **category wins**: the trustee selected by the means of category is better than the one selected with individual recommendation; then this method gets one point;

- **equal result**: if the difference between the two trustworthiness values is not enough (it is under a threshold), we consider it as indistinguishable result. In particular, we considered the threshold of 3%.

These two phases are repeated 500 times.

3.4 Outputs

In every simulation we use some different indexes to analyze its results:

1. **trustee wins**: number of times in which the trustee selected with individual recommendation is better than the one selected by the means of categorial recommendation;
2. **category wins**: number of times in which the trustee selected by the means of categorial recommendation (the nearest agent belonging to it) is better than the one selected with individual recommendation;
3. **equal result**: number of times in which the difference between the two trustworthiness values is less than 3%;
4. **trustee mean**: average value of trustees' trustworthiness chosen with individual recommendation in the 500 run;
5. **category mean**: average value of the trustees' trustworthiness chosen with the categorial recommendation in the 500 run.

4. SIMULATIONS RESULT

In these simulations we present a series of scenarios with different settings to show when it is more convenient to exploit recommendations about categories rather than recommendations about individuals, and vice versa.

We also present the "all-in-one" scenario, whose peculiarity is that the exploration lasts just 1 tick and in that tick every trustee experiences all the others. Although this is a limit case, very unlikely in the real world, it is really interesting as each trustee has not a good knowledge of the other trustees as individual elements (it has experienced them just one time), but it is able to get a really good knowledge of their categories, as it has experienced them as many times as the number of trustees for each category. So this is an explicit case in which the recommendations of the trustees about categories are surely more informative than the ones about individuals.

Simulations' results are presented in a tabular and graphical way. In particular, we have chosen to highlight in tables, with a yellow color, cases in which category's performance overtakes or equalizes individual's one.

4.1 First Simulation

In this first set of simulations we use 20 trustees for category and analyze what happens when both the duration of exploration phase and the percentage of queried trustees change.

Tables' legend:

- : cases in which category's performance overtakes or equalizes individual's one.

First scenario:

- Trustees queried by the trustor: 100%

Table 1. 80 trustees, 100% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	324	42	134	0,796	0,854
50	252	93	155	0,799	0,831
25	226	140	134	0,802	0,811
10	184	179	137	0,800	0,785
5	189	191	120	0,780	0,756
3	158	227	115	0,781	0,729
1	133	289	78	0,754	0,649
all-in-one	118	266	116	0,8	0,727

Second scenario:

- Trustees queried by the trustor: 50%

Table 2. 80 trustees, 50% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	277	73	150	0,799	0,841
50	227	127	146	0,801	0,811
25	182	170	148	0,796	0,782
10	176	210	114	0,778	0,739
5	159	225	116	0,763	0,702
3	150	243	107	0,749	0,684
1	145	280	75	0,723	0,618
all-in-one	94	313	93	0,803	0,689

Third scenario:

- Trustees queried by the trustor: 25%

Table 3. 80 trustees, 25% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	248	113	139	0,803	0,824
50	218	158	124	0,790	0,787
25	193	192	115	0,779	0,755
10	159	222	119	0,756	0,705
5	160	244	96	0,717	0,651
3	145	264	91	0,712	0,637
1	169	255	76	0,667	0,587
all-in-one	83	336	81	0,803	0,656

Fourth scenario:

- Trustees queried by the trustor: 10%

Table 4. 80 trustees, 10% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	209	156	135	0,794	0,784
50	184	197	119	0,774	0,742
25	159	248	93	0,754	0,685
10	175	230	95	0,691	0,642
5	176	241	83	0,671	0,614
3	169	247	84	0,661	0,600
1	170	259	71	0,615	0,548
all-in-one	83	346	71	0,796	0,619

Fifth scenario:

- Trustees queried by the trustor: 5%

Table 5. 80 trustees, 5% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	189	184	127	0,772	0,757
50	188	225	87	0,751	0,709
25	174	220	106	0,700	0,649
10	188	219	93	0,659	0,616
5	174	228	98	0,648	0,611
3	176	248	76	0,637	0,589
1	190	235	75	0,603	0,559
all-in-one	91	337	72	0,769	0,606

Below we synthetize these results in two graph (one for the "t win" dimension and the other for the "c win" dimension).

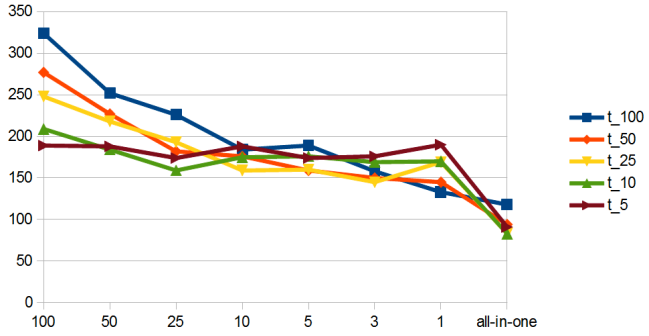

Figure 1. Trustee wins when there are 20 trustees for category.

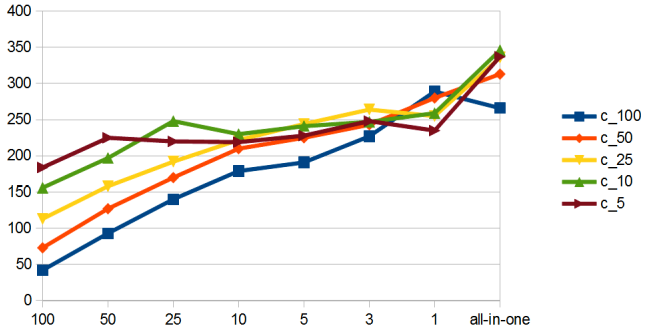

Figure 2. Category wins when there are 20 trustees for category.

In the first graph it is easy to see how the value of "trustee wins" decreases when decreases the number of ticks in the exploratory phase, that is when is reduced the number of interactions among the agents before being queried; on the contrary, the value of "category wins" increases proportionally with this reduction (*first effect*).

At the same time, there is a direct proportionality between the value of "trustee wins" and the number of trustees queried in the querying phase; while the value of "category wins" increases proportionally with the reduction of the number of trustees queried (*second effect*).

In practice, both these effects seem suggest how the role of categories becomes relevant when either decreases and degrades the knowledge within the analyzed system (before the interaction with the trustor) or is reduced the transferred knowledge (to the trustor).

Let us explain better. The *first effect* can be described with the fact that each agent, reducing the number of interactions with the other agents in the explorative phase, will have relevantly less information with respect to the individual agents. At the same time its knowledge with respect to categories does not undergo a significant decline given that categories' performances derive from several different agents.

The *second effect* can be explained with the fact that reducing the number of queried trustees, the trustor will receive with decreasing probability information about the more trustworthy individual agents in the domain, while information on categories, maintains a good level of stability also reducing the number of queried agents, thanks to greater robustness of these structures.

Resuming, the above pictures clearly show how, when the quantity of information (about the agents' trustworthiness exchanged in the system) decreases, it is better to rely on the categorial recommendations rather than individual recommendations.

This result reaches the point of highest criticality in the "all-in-one" case in which, as expected, "trustee wins" returns the minimal value and "category wins" returns the maximal value.

4.2 Second Simulation

In the second set of simulations we try to increment the number of trustees to 100 for category. It means that each trustee has much more neighbors than before.

Tables' legend:

- : cases in which category's performance overtakes or equalizes individual's one.

Sixth scenario:

- Trustees queried by the trustor: 100%

Table 6. 400 trustees, 100% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	401	5	94	0,796	0,882
50	382	8	110	0,803	0,879
25	372	15	113	0,802	0,873
10	319	43	138	0,802	0,86
5	323	53	124	0,795	0,85
3	271	95	134	0,801	0,834
1	151	238	111	0,803	0,759
all-in-one	155	252	93	0,796	0,741

Seventh scenario:

- Trustees queried by the trustor: 50%

Table 7. 100 trustees, 50% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	382	3	115	0,799	0,878
50	362	27	111	0,801	0,873
25	354	20	126	0,799	0,866
10	292	65	143	0,804	0,849
5	291	79	130	0,8	0,842
3	221	142	137	0,803	0,813
1	120	276	104	0,797	0,712
all-in-one	139	270	91	0,800	0,727

Eighth scenario:

- Trustees queried by the trustor: 25%

Table 8. 100 trustees, 25% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	367	14	119	0,797	0,872
50	356	29	115	0,799	0,867
25	351	39	110	0,798	0,859
10	273	100	127	0,803	0,836
5	276	102	122	0,795	0,83
3	212	144	144	0,801	0,802
1	113	289	98	0,801	0,705
all-in-one	130	274	96	0,797	0,702

Ninth scenario:

- Trustees queried by the trustor: 10%

Table 9. 100 trustees, 10% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	344	26	130	0,797	0,864
50	318	48	134	0,802	0,858
25	271	84	145	0,804	0,843
10	239	122	139	0,802	0,82
5	223	151	126	0,797	8,02
3	176	217	107	0,803	0,769
1	106	318	76	0,793	0,663
all-in-one	92	322	86	0,796	0,654

Tenth scenario:

- Trustees queried by the trustor: 5%

Table 10. 100 trustees, 5% queried by the trustor

Expl. Ph.	T win	C win	Equal	C Av	T Av
100	316	47	137	0,802	0,856
50	310	68	122	0,797	0,842
25	262	100	138	0,798	0,823
10	205	159	136	0,801	0,799
5	190	197	113	0,8	0,772
3	133	257	110	0,8	0,725
1	99	335	66	0,792	0,63
all-in-one	80	353	67	0,802	0,622

Again, we summarize the results into two graph.

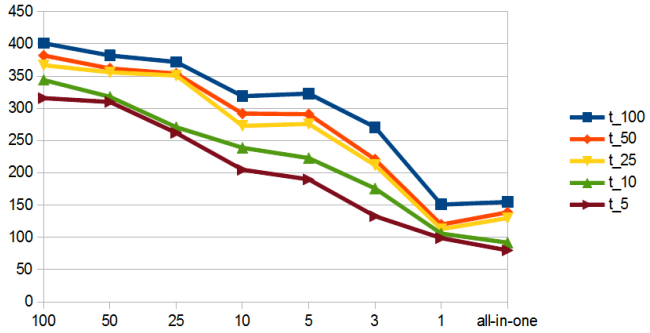

Figure 3. Trustee wins when there are 100 trustees for category.

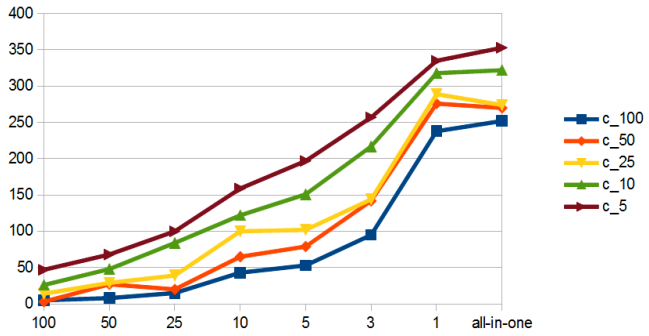

Figure 4. Category wins when there are 100 trustees for category.

In this second set of simulations, are confirmed the two effects detected in the first simulations. However it is possible observe a greater difficulty of recommendations about categories to prevail on the recommendations about individuals: just strongly reducing the trustees queried by the trustor it is possible value a role for categories' recommendations.

This result could be explained with the fact that increasing the number of the agents in the neighborhood of each agent, it increases the possibility to have in it highly trustworthy agents and as a consequence more agents reporting information about them.

5. CONCLUSIONS

In other works [9][10][2] were shown the advantages of using reasoning about categorization for selecting trustworthy agents. In particular, how it were possible to attribute to a certain unknown agent, a value of trustworthiness with respect to a specific task, on the basis of its classification in, and membership to, one (/or more) category/ies. In practice, the role of generalized knowledge and prejudice (in the sense of pre-established judgment on the agents belonging to that category) has proven to determine the possibility to anticipate the value of unknown agents.

In this paper we have investigated the different roles that can play recommendations about individual agents and about categories of agents.

In this case the new agent introduced (called trustror) has a whole world of agents completely unknown to it, and ask for recommendations to a (variable) subset of agents for selecting an agent to whom delegate a task. The information received regards both individual agents and agents' categories. The informative power of these two kinds of recommendations is dependent from the previous interactions among the agents and also from the number agents queried by the trustor. However, there are cases in which information about categories is more useful that information towards individual agents. In some sense this result complements the results achieved in [9][10][2] because here we have a more strict match between information on individual agents and information about categories of agents: We are measuring the quantity of information, about individual agents and categories, for evaluating when is better using *direct information* rather than *generalized information* or, vice versa, when is better using the positive power of prejudice. Our results show how in certain cases becomes essential the use of categorial knowledge for selecting qualified partners.

In this work we have in fact considered a closed world, with a fixed set of agents. This choice was based on the fact that we were interested to evaluate the relationships between knowledge about individual and knowledge about categories, for calibrating their roles and reciprocal influences. In future works we have to consider how, starting from the analysis of this study, could change the role of knowledge about categories in a situation of open world. In particular, we could experiment the dynamic of this role with respect to the stability of the performances of the different agents becoming to a category.

6. REFERENCES

[1] Adomavicius, G., Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering (TKDE) 17, 734–749, 2005

[2] Burnett, C., Norman, T., and Sycara, K. 2010. Bootstrapping trust evaluations through stereotypes. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10). 241248.

[3] C. Burnett, T. J. Norman, and K. Sycara. Stereotypical trust and bias in dynamic multiagent systems. ACM Transactions on Intelligent Systems and Technology (TIST), 4(2):26, 2013.

[4] Castelfranchi C., Falcone R., Trust Theory: A Socio-Cognitive and Computational Model, John Wiley and Sons, April 2010.

[5] Conte R., and Paolucci M., 2002, Reputation in artificial societies. Social beliefs for social order. Boston: Kluwer Academic Publishers.

[6] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Improving Recommendation Quality by Merging Collaborative Filtering and Social Relationships. In Proc. of the International Conference on Intelligent Systems Design and Applications (ISDA 2011) , Córdoba, Spain, IEEE Computer Society Press, 2011

[7] *P De Meo, E Ferrara, D Rosaci, and G Sarné*. Trust and Compactness of Social Network Groups**.** IEEE Transactions on Cybernetics, PP:99, 2014

[8] Falcone R., Castelfranchi C. *Generalizing Trust: Inferencing Trustworthiness from Categories*. In: TRUST 2008 - Trust in Agent Societies, 11th International Workshop, TRUST 2008. Revised Selected and Invited Papers (Estoril, Portugal, 12-13 May 2008). Proceedings, pp. 65 - 80. R. Falcone, S. K. Barber, J. Sabater-Mir, M. P. Singh (eds.). (Lecture Notes in Artificial Intelligence, vol. 5396). Springer, 2008.

[9] Falcone R., Piunti, M., Venanzi, M., Castelfranchi C., (2013), From Manifesta to Krypta: The Relevance of Categories for Trusting Others, in R. Falcone and M. Singh (Eds.) Trust in Multiagent Systems, ACM Transaction on Intelligent Systems and Technology, Volume 4 Issue 2, March 2013

[10] Falcone R., Sapienza A., Castelfranchi C.,The relevance of Categories for trusting Information Sources,"Transactions on Internet Technology", submitted

[11] H. Fang, J. Zhang, M. Sensoy, and N. M. Thalmann. A generalized stereotypical trust model. In Proceedings of the 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pages 698–705, 2012.

[12] G. Guo, J. Zhang and N. Yorke-Smith, Leveraging Multiviews of Trust and Similarity to Enhance Clustering-

based Recommender Systems, Knowledge-Based Systems, accepted, 2014

[13] Huynh, T.D., Jennings, N. R. and Shadbolt, N.R. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems*, 13, (2), 119-154., 2006

[14] P. Lops, M. Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook. Springer, pp. 73–105, 2011.

[15] P. Massa, P. Avesani, Trust-aware recommender systems, RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems, 2007

[16] S. Ramchurn, N. Jennings, Carles Sierra, and Lluis Godo. Devising a trust model for multi-agent interactions using confidence and reputation. Applied Artificial Intelligence, 18(9-10):833-852, 2004.

[17] Sabater-Mir, J. 2003. Trust and reputation for agent societies. Ph.D. thesis, Universitat Autonoma de Barcelona.

[18] M. Sensoy, B. Yilmaz, and T. J. Norman. STAGE: Stereotypical trust assessment through graph extraction. Computational Intelligence, 2014.

[19] C. Than and S. Han, Improving Recommender Systems by Incorporating Similarity, Trust and Reputation, Journal of Internet Services and Information Security (JISIS), volume: 4, number: 1, pp. 64-76, 2014

[20] Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

[21] Yolum, P. and Singh, M. P. 2003. Emergent properties of referral systems. In Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS'03).