Reputation for Innovating Social Networks*

Rosaria Conte[1], Mario Paolucci[1], and Jordi Sabater-Mir[2]

[1] Institute for Cognitive Science and Technology
Via San Martino della Battaglia, 44, Rome, ITALY
{rosaria.conte|mario.paolucci}@istc.cnr.it
[2] Artificial Intelligence Research Institute, Barcelona, SPAIN
jsabater@iiia.csic.es

Abstract. Reputation is a fundamental instrument of partner selection. Developed within the domain of electronic auctions, reputation technology is being been imported into other applications, from social networks to institutional evaluation. Its impact on trust enforcement is uncontroversial and its management is of primary concern for entrepreneurs and other economic operators.
In the present paper, we will shortly report upon simulation-based studies on the role of reputation as a more tolerant form of social capital than familiarity networks. Whereas the latter exclude non-trustworthy partners, reputation is a more inclusive mechanism upon which larger and more dynamic networks are constructed. After the presentation of the theory of reputation developed by the authors in the last decade, a computational system (REPAGE) for forming and exchanging reputation information will be presented and findings from experimental simulations recently run on this system will be resumed. Final remarks and ideas for future works will conclude the paper.
Keywords: Artificial societies, Reputation, Innovation, Social Networks

1 The Problem

In marketplaces, and more generally in social exchange, reputation provides traders and other users with a fundamental instrument of partner selection. Developed within the domain of electronic auctions (like eBay, cf. for a survey [**?**]), in the last few years reputation technology has been invading other electronic applications, from social networks to institutional evaluation. Its impact on trust enforcement is so uncontroversial, that corporate reputation is counted as an asset, and its management is of primary concern for entrepreneurs and other economic operators [**?**]. Nowadays, one can make money by assisting people in dealing with, managing, and even refreshing their own reputation[3]. Such

* This work was partially supported by the European Community under the FP6 programme (*eRep* project, contract number CIT5-028575) and by the Italian Ministry of University and Scientific Research under the FIRB programme (Socrate project, contract number RBNE03Y338)
[3] cf. http://www.reputationdefender.com

a far-reaching confidence in reputation probably rests on the assumption that it supports us in the complimentary roles of selecting trustable partners and being selected as such.

Far from discrediting the view of reputation as a trust enhancement mechanism, we would like to enlarge the boundaries of the phenomenon at stake, by pointing to another functionality, namely the enlargement and innovation of social networks.

The rest of the paper will unfold as follows. First, the role of image-based networks in a world where the boundaries of social and trading networks are constantly widened will be questioned. Next, drawing upon the social cognitive model presented in [**?**], a notion of reputation as a special form of social evaluation will be re-proposed. This notion will be argued to allow for network innovation: on one hand, reputation allows for social evaluation to circulate and complement one's personal experience. On the other, it will be argued to accomplish a most crucial and delicate task, i.e. check and discard misinformation without necessarily discarding the agents responsible for its transmission. In other words, reputation networks will be shown to be more inclusive than image-networks ceteris paribus, and at the same time to help checking the truth-value of the information circulating in the network.

2 Main Claim and Organization of the Paper

The paper is aimed to discuss the view of reputation in the framework presented above. It builds upon the state of the art on reputation theory and technology at the Laboratory of Agent Based Social Simulation (LABSS) of the Institute of Cognitive Science and Technology (ISTC), within the *eRep* project[4] . The starting point will be the results from experimental simulations presented in [**?**], thanks to the computational system REPAGE, worked out by the authors and presented in [**?**].

In [**?**], experiments were meant to show the value added of reputation as a mechanism of partner selection. Results show that an artificial market where agents exchange both image and reputation obtains better results in terms of production quality than a market were agents exchange their own opinions about one another. The reason for such a difference lies in retaliation: as will be argued later on in the present paper, image-based, or familiarity, networks perform more poorly than reputation networks exactly because they induce retaliation.

In the present paper, we will shortly report upon previous findings in order to put forward a more general hypothesis, which seems to be supported by our simulations. Reputation allows for a far more tolerant, gross-grained social selector than image. Hence, whereas shared image forms a selective platform on which familiarity networks that exclude non-trustworthy partners are constructed, reputation is a rather more inclusive mechanism upon which larger and more dynamic networks are constructed. Thanks to it,

[4] http://megatron.iiia.csic.es/eRep/

– candidate (non-confirmed) information may circulate allowing the network to learn new social knowledge,
– the network may innovate, by integrating new partners,
– and put up with errors without discarding the partners that fell prey to them.

In a few words, reputation appears as a more dynamic form of social capital, allowing for social networks to be innovated.

The paper is organized as follows: after the synthetic presentation of the theory of reputation developed by the authors, the REPAGE system will be presented and the experimental simulation recently run by the authors thanks to such a system will be resumed. The findings from that study will be rediscussed in the light of the present hypothesis. Final remarks and ideas for future works will conclude the paper.

3 A Social Cognitive Model of Reputation

In this section we will report on a social cognitive model of reputation presented in [**?**], where

– the difference between image and reputation has been introduced,
– the different roles agents play when evaluating someone and transmitting this evaluation are analysed,
– the decision processes based upon both image and reputation are examined.

A cognitive process involves symbolic mental representations (such as goals and beliefs) and is effectuated by means of the mental operations that agents perform upon these representations (reasoning, decision-making, etc.). A social cognitive process is a process that involves social beliefs and goals, and that is effectuated by means of the operations that agents perform upon social beliefs and goals (e.g., social reasoning). A belief or a goal is social when it mentions another agent and possibly one or more of his or her mental states (for a discussion of these notions, see [**?**], [**?**]).

The social cognitive approach is receiving growing attention within several subfields of the Sciences of the Artificial, in particular intelligent software agents, Multi-Agent Systems, and Artificial Societies. Unlike the "theory of mind" (cf. [**?**]) approach, this approach aims at modelling and possibly implementing systems acting in a social (whether natural or artificial) environment. The theory of mind focuses upon one aspect, although an important one, of social agency, i.e., social beliefs (knowledge agents have about others).

Here, the approach adopted is aimed at modelling the variety of mental states (including social goals, motivations, obligations) and operations (such as social reasoning and decision-making) necessary for an intelligent social system to act in some domain and influence other agents (social learning, influence, and control).

3.1 Image and Reputation

The social cognitive model is a dynamic approach that considers reputation as the output of a social process of transmission of information. The input to this process is the evaluation that agents directly form about a given agent during interaction or observation. This evaluation will be called here the social image of the agent. An agent's reputation is argued to be distinct from, although strictly interrelated with, its image. More precisely, image will be defined as a set of evaluative beliefs about a given target, while reputation will be defined as the process and the effect of transmission of image. As an application of this model, some simple predictions made possible by this conceptualisation will be presented. Furthermore, the decision to accept image will be compared with and distinguished from the decision to acknowledge reputation. Image consists of a set of evaluative beliefs [**?**] about the characteristics of the target, i.e. it is an assessment of its positive or negative qualities with regard to a norm, a competence, and so on.

Reputation is both the process and the effect of transmission of a target's image. The image relevant for social reputation may concern a subset of the target's characteristics, i.e., its willingness to comply with socially accepted norms and customs. More precisely, reputation is defined to consist of three distinct but interrelated objects:

- a cognitive representation, or more precisely a believed evaluation;
- a population object, i.e., a propagating believed evaluation;
- an objective emergent property at the agent level, i.e., what the agent is believed to be.

In fact, reputation is a highly dynamic phenomenon in two distinct senses: it is subject to change, especially as an effect of corruption, errors, deception, etc.; and it emerges as an effect of a multi-level bidirectional process. In particular, it proceeds from the level of individual cognition to the level of social propagation and from this level back to that of individual cognition again. What is more interesting, once it gets to the population level, it gives rise to a further property at the agent level: agents acquire a bad or good name. Reputation is not only what people think about targets but also what targets are in the eyes of others. From the very moment agents are targeted by the community, want it or not and believe it or not, their lives change: reputation becomes the immaterial, more powerful equivalent of a scarlet letter sewed to their clothes. It is more powerful because it may not even be perceived by those to whom it sticks, and consequently it is out of their control. Reputation is an objective social property that emerges from a propagating cognitive representation, which lacks an identified source, whereas image always requires that at least one evaluator to be identified.

3.2 Reputation and Image As Social Evaluations

According to [**?**], an evaluation is a hybrid representation. An agent has an evaluation when he or she believes that a given entity is good for, or can achieve, a

given goal. An agent has a social evaluation when his or her belief concerns another agent as a means for achieving this goal. A given social evaluation includes three sets of agents:

- a nonempty set E of agents who share the evaluation (evaluators)
- a nonempty set T of evaluation targets
- a nonempty set B of beneficiaries, i.e., the agents sharing the goal with regard to which the elements of T are evaluated.

Often, evaluators and beneficiaries coincide, or at least have nonempty intersection but this is not necessarily the case. A given agent t is a target of a social evaluation when t is believed to be a good/bad means for a given goal of the set of agents B, which may include or not the evaluator. (Social) evaluations may concern physical, mental, and social properties of targets; agents may evaluate a target as to both its capacity and willingness to achieve a shared goal. In particular, more or less explicitly, social evaluations concern the targets' willingness to achieve a goal or interest. Formally, e (with $e \in E$) may evaluate t (where $t \in T$) with regard to a state of the world that is in b 's (with $b \in B$) interest, but of which b may not be aware.

The interest/goal with regard to which t is evaluated may be a distributed or collective advantage. It is an advantage for the individual members who are included in the set B, or it may favour a supra individual entity, which results from interactions among the members of B (for example, if B's members form a team).

It is very easy to find social examples where the three sets coincide: universal norms, such as "Don't commit murder," apply to, benefit, and get evaluated from the whole universe of agents.

There are situations in which beneficiaries, targets, and evaluators are separated, for example, when norms safeguard the interests of a subset of the population. Consider the quality of TV programs during the children's timeshare. Here, we can find three clearly separated sets: children are the beneficiaries, while the adults entrusted with taking care of the children are the evaluators. Of course, here the intersection between B and E still exists, because E may be said to adopt B's interests. But who are the targets of evaluation? Not all the adults, but the writers of programs and the decision-makers at the broadcast stations. In this case, there is a nonempty intersection between E and T but no full overlap. Also, if the target of evaluation is the broadcaster itself, a supra-individual entity, then the intersection can be considered to be null: $E \cap T = \emptyset$.

To assume that a target t is assigned a given reputation implies assuming that t is believed to be "good" or "bad," but it does not imply sharing either evaluation. Reputation then involves four sets of agents:

- a nonempty set E of agents who share the evaluation
- a nonempty set T of evaluation targets
- a nonempty set B of beneficiaries, i.e., the agents sharing the goal with regard to which the elements of T are evaluated

– a nonempty set M of agents who share the meta-belief that members of E share the evaluation; this is the set of all agents aware of the effect of reputation (as stated above, effect is only one component of it; awareness of the process is not implied).

Often, E can be taken as a subset of M; the evaluators are aware of the effect of evaluation. In most situations, the intersection between the two sets is at least nonempty, but exceptions exist. M in substance is the set of reputation transmitters, or third parties. Third parties share a meta-belief about a given target, whether they share the concerned belief or not. In real matters, agents may play more than one role simultaneously.

3.3 Reputation-Based Decisions

The model presented above focuses on the definition of some critical sets, defining characteristics that we believe to be relevant for reputation. On the basis of our definitions, we will go on from examining the main decision processes undertaken by social agents with regard to image and reputation. To understand the difference between image and reputation, the mental decisions based upon them must be analysed at the following three levels:

– Epistemic: accept the beliefs that form either a given image or acknowledge a given reputation. This implies that a believed evaluation gives rise to one's direct evaluation. Suppose I know that the friend I mostly admire has a good opinion of Mr. Bush. However puzzled by this dissonance-inducing news, I may be convinced by my friend to accept this evaluation and share it.
– Pragmatic - Strategic: use image in order to decide whether and how to interact with the target. Once I have my own opinion (perhaps resulting from acceptance of others' evaluations) about a target, I will use it to make decisions about my future actions concerning that target. Perhaps, I may abstain from participating in political activity against Mr. Bush.
– Memetic: transmit my (or others') evaluative beliefs about a given target to others. Whether or not I act in conformity with a propagating evaluation, I may decide to spread the news to others. Image and reputation are distinct objects. Both are social in two senses: they concern another agent's (the target) properties (the target's presumed attitude towards socially desirable behaviour), and they may be shared by a multitude of agents. However, the two notions operate at different levels. Image is a belief, namely, an evaluation. Reputation is a meta-belief, i.e., a belief about others' evaluations of the target with regard to a socially desirable behaviour.

The epistemic decision level is grounded upon both image and reputation. An epistemic decision concerns whether to accept a given belief. In the case of image, it concerns evaluations; in the case of reputation, it concerns meta-beliefs (others' evaluations). Both these decisions are relatively independent of one another. To accept a meta-belief does not require that the first-level belief

be held to be true, and viceversa: to accept a given image about someone does not imply a belief that that person enjoys the corresponding reputation. To accept/form a given image about a target implies an assessment of the truth value of evaluations concerning the target. In contrast, reputation consists of meta- beliefs about image, i.e., about others' evaluative beliefs concerning the holder.

Conversely, to acknowledge a given reputation does not lead to sharing others' evaluations but rather to the belief that these evaluations are held or circulated by others. To assess the value of such a meta-belief is a rather straightforward operation. For the recipient to be relatively confident about this meta-belief, it is probably sufficient that it be exposed to rumours. In order to understand the difference between image acceptance and reputation acknowledgement, it is necessary to investigate the different roles of image and reputation beliefs in the agents' minds.

But before setting out to do so, a couple of intertwined preliminary conclusions can be suggested. First, reputation is less likely to be falsified than image. Second, the process of transmission, rather than its effect, is prevalent in reputation. In fact, it is more difficult to ascertain whether a given state is true in anyone's mind than in the external world. An external state of the world is more controllable than a mental one. It is relatively difficult to check whether, to what extent, and by whom that state of the world is believed to be true. But the representation of another's mental state is essential for social reasoning, and any clue to such a belief, given a lack of other indications, is better than no information. This easy acceptance of reputation information gives prevalence to the process over the content. Therefore, any study on reputation that concentrates on content only is likely to miss the point completely.

Agents resort to their evaluative beliefs in order to achieve their goals [**?**]. Evaluations are guidelines for planning; evaluations about other agents are guidelines for social action and social planning. Therefore, the image a given agent has about t will guide its action wrt t, will suggest whether it is convenient to interact with t or not, and will also suggest what type of interaction to establish with t. Of course, image may be conveyed to others in order to guide their actions towards the target in a positive or negative sense. When transmitting its image of t, the agent attempts to influence others' strategic decisions. To do so, the agent must (pretend to) be committed to the evaluation and take responsibility for its truth value before the recipient. Reputation enters direct pragmatic or strategic decisions when it is consistent with image or when no image of the target has been formed. Otherwise, in pragmatic or strategic decisions, reputation is often superseded by image. However, in influencing others' decisions, the opposite pattern occurs: in this case, only reputation considerations apply. Agents tend to influence others' social decisions by transmitting to them information about the target's reputation. Two main reasons explain this inverse pattern:

– agents expect that a general opinion, or at least a general voice, is more credible and acceptable than an individual one

– agents reporting on reputation do not need to commit to its truth value, and do not take responsibility over it; consequently, they may influence others to a lower personal cost.

The memetic decision can be roughly described as the decision to spread reputation. In the case of communication about reputation the communicative action is performed in order to

– obtain the goal that the hearer believes that t is assigned a given reputation by others, rather than by the speaker himself of herself (g2), and to
– obtain the goal that the hearer propagates t's reputation (g4), possibly but not necessarily by having him believe that t is in fact assigned a given reputation (g3).

Whilst g2 is communicative (the speaker wants the hearer to believe that the speaker used the language to achieve that effect), g4 is not. (Indeed, the speaker usually conceals this intention under the opposite communication: *I tell you in confidence, therefore don't spread the news....*)

Consequently, communication about reputation is a communication about a meta-belief, i.e., about others' mental attitudes. To spread news about someone's reputation does not bind the speaker to commit himself to the truth value of the evaluation conveyed but only to the existence of rumours about it. Unlike ordinary sincere communication, only the acceptance of a meta-belief is required in communication about reputation. And unlike ordinary deception (for a definition of the latter, see [**?**]), communication about reputation implies

– no personal commitment of the speaker with regard to the main content of the information delivered. If speaker reports on t's bad reputation, he is by no means stating that t deserved it; and
– no responsibility with regard to the credibility of (the source of) information (*I was told that t is a bad guy*).

Two points ought to be considered here. First, the source of the meta-belief is implicit (*I was told...*). Secondly, the set of agents to whom the belief p is attributed is non-defined (*t is ill/well reputed*). Of course, the above points do does not mean that communication about reputation is always sincere. Quite on the contrary, one can and does often deceive about others' reputation. But to be effective the liar neither commits to the truth of the information transmitted nor takes responsibility with regard to its consequences. If one wants to deceive another about reputation, one should report it as a rumour independent of or even despite one's own beliefs!

4 The Antisocial Effects of Image

The model points to several consequences of image (I) and reputation (R) spreading. Let us examine them with some detail.

First, both I and R spreading are forms of cooperation. Both provide the cognitive matter to informational reciprocity, allowing for material cooperation to take place: agents exchanging shared information about whom they believe to be good and whom they believe to be bad in the group, market, organization or society cooperate at the level of information. By doing so, they allow for material reciprocators, good sellers, norm observers and other good guys to survive and compete with cheaters. Hence, both image and reputation lead to material cooperation.

Secondly, both are expected to lead to social cohesion. Obviously, cheaters may bluff and try to play as informational reciprocators in order to enjoy the benefits of a good image without sustaining the costs of acquiring one. But once bluff is found out, stable social sub-nets are formed by reliable informers who will be sitting there as long as possible. These subnets are more or less what economists and other social scientists call familiarity networks, characterized by reciprocal acquaintance, even benevolence, and trust.

Third, and consequently, both I and R are expected to lead to a reduction in the dimensions of the network of material cooperation or exchange. Acting as selectors, they lead to the initial set of potential relationships to be reduced. Here is where the difference between I and R starts to emerge. I is more selective and R is more inclusive. What is more, unlike R, I spreading reveals the identity of evaluators, or of a subset of them. Shared evaluations make the sources vulnerable, exposing them to possible retaliations. Instead, reported on evaluations protect the identities of evaluators, discouraging or preventing retaliations.

Of course, reported on evaluations provide only candidate evaluations, which often turns to be false and therefore useless. However, one can argue that to find a R disconfirmed is less disruptive than I being disconfirmed. When finding an I received by someone to be wrong, the recipient will face a rather distressing alternative: the source is either misinformed or ill-willed. Either information is unreliable, or the informer's intention is wicked. In any case, the informer cannot be trusted any more, and must be set apart if not punished. Hence, the disruptive effect of image spreading is a function of the amount of informational error and cheating injected into the network. An image-based social network is expected to be rigid, meaning rather sensitive to errors: if a given threshold of error is overcome, the whole system is probably bound to fall apart, and the network will be fatally affected by distrust.

The reason for expecting such a gloomy perspective is complex. For one thing, once recipients of false image have reacted negatively, either getting rid of their bad informers or taking their revenge against them, balance is hardly restored. Mutual defeat will not stop so easily, and retaliation will tend to call for further retaliation in a chain of self-fulfilling prophecies that is usually fatal on both sides. In a stock market, this may even turn into a general collapse.

With reputation, instead, the quality of information received is not necessarily nor immediately tested before being passed on. Misinformation may not be found out so soon, and even when it is finally disclosed, it will not lead the recipient to question the quality of the informer, simply because the latter never

committed itself to the truth value of the information conveyed. The reputation network is expected to be more robust than the image-based one, as it puts up with a far larger amount of misinformation without discarding nor punishing the vectors of misinformation, which in fact are not always responsible for such errors.

In the rest of the paper, we will see whether such expectations are met by existing simulation evidence. This was gathered in a study by [**?**], where our system REPAGE - a REPutation and imAGE tool developed on the grounds of the theory of reputation - was implemented on an agent architecture in order to reproduce an artificial market. In such a setting, buyers were allowed to use either image only (L1 condition) or image plus reputation (L2 condition), and the effects of these two settings were compared in terms of averaged and accumulated quality of products. After a short description of REPAGE, we will turn to show the relevance of these artificial findings to the present view of image and reputation.

5 Repage Model and Architecture

Repage [**?**] is a computational system based on the theory of reputation presented above [**?**]. Its architecture includes three main elements, a memory, a set of detectors and the analyzer.

The memory is composed by a set of references to the predicates hold in the main memory of the agent. Predicates are conceptually organized in levels and inter-connected. Each predicate that belongs to one of the main types (including image and reputation) contains a probabilistic evaluation that refers to a certain agent in a specific role. For instance, an agent may have an image of agent T (target) as a seller (role), and a different image of the same agent T as informant. The probabilistic evaluation consist of a probability distribution over the discrete sorted set of labels: Very Bad, Bad, Normal, Good, Very Good. The network of dependences specifies which predicates contribute to the values of others. In this sense, each predicate has a set of precedents and a set of antecedents.

The detectors, inference units specialized in each particular kind of predicate, receive notifications from predicates that change or that appear in the system and use dependencies to recalculate the new values or to populate the memory with new predicates. Each predicate has associated a strength that is function of its antecedents and of the intrinsic properties of each kind of predicate. As a general rule, predicates that resume or aggregate a larger number of predicates will hold a higher strength.

At the first level of the Repage memory we find a set of predicates not evaluated yet by the system. Contracts are agreements on the future interaction between two agents. Their result is represented by a Fulfillment. Communications is information that other agents may convey, and may be related to three different aspects: the image that the informer has about a target, the image that, according to the informer, a third party agent has on the target, and the reputation that the informer has about the target.

In level two we have two kinds of predicates. Valued communication is the subjective evaluation of the communication received that takes into account, for instance, the image the agent may have of the informer as informant. Communications from agents whose credibility is low will not be considered as strong as the ones coming from well reputed informers. An outcome is the agent's subjective evaluation of a direct interaction, built up from a fulfillment and a contract. At the third level we find two predicates that are only fed by valued communications. On one hand, a shared voice will hold the information received about the same target and same role coming from communicated reputations. On the other hand, shared evaluation is the equivalent for communicated images and third party images.

Shared voice predicates will finally generate candidate reputation; shared evaluation together with outcomes will generate candidate image. Newly generated candidate reputation and image are usually not strong enough; new communications and new direct interactions will contribute to reinforce them until a threshold, over which they become full-fledged image or reputation. We refer to [**?**] for a much more detailed presentation. From the point of view of the agent strucuture, integration with the other parts of our deliberative agents is stright-forward. Repage memory links to the main memory of the agent that is fed by its communication and decision making module, and at the same time, this last module, the one that contain all the reasoning procedures uses the predicates generated by Repage to make decisions.

6 Simulation Experiment

In [**?**] we applied our system REPAGE to a simulation experiment of the simplest setting in which accurate information is a commodity: an agent-based market with instability. The model has been designed with the purpose of providing the simplest possible setting where information is both valuable and scarce. The system must be considered as a proof of concept, not grounded on micro or macro data, but providing an abstract economic metaphor. This simplified approach is largely used in the reputation field (see for example [**?**]), both on the side of the market design and of the agent design in the study of market with asymmetric information. We follow this approach since our main interest is on the side of agent design, and we must be able to clearly separate complex effect due to agent structure from ones due to market structure.

6.1 Design of the Experiment

The experiment includes only two kind of agents, the buyers and the sellers. All agents perform actions in discrete time units (turns from now on). In a turn, a buyer performs one communication request and one purchase operation. In addition, the buyer answers all the information requests that it receives. Goods are characterized by an utility factor that we interpret as quality (but, given the

level of abstraction used, could as well represent other utility factors as quantity, discount, timeliness) with values between 1 and 100.

Sellers are characterized by a constant quality, drawn following a stationary probablilty distribution, and a fixed stock, that is decreased at every purchase; they are essentially reactive, their functional role in the simulation being limited to providing an abstract good of variable quality to the buyers. Sellers exit the simulation when the stock is exhausted and are substituted by a new seller with similar characteristics but with a new identity (and as such, unknown to the buyers). This continuous seller update characterises our model, for example in comparison with recent work as [**?**], where both sellers and buyers are essentially fixed.

The disappearance of sellers makes information necessary; reliable communication allows for faster discover of the better sellers. This motivates the agents to participate in the information exchange. In a setting with permanent sellers (infinite stock), once all buyers have found a good seller, there is no reason to change and the experiment freezes. With finite stock, even after having found a good seller, buyers, should be prepared to start a new search when the good seller's stock ends.

At the same time, limited stock makes good sellers a scarce resource, and this constitutes a motivation for the agents not to distribute information. One of the interests of the model is in the balance between these two factors.

There are four parameters that describe an experiment: the number of buyers NB, the number of sellers NS, the stock for each seller S, and the distribution of quality among sellers. We defined the two main experimental situations, L1 where there is only exchange of image, and L2 where both image and reputation are used.

6.2 Decision Making Module

In [**?**], the decision making procedure was shown to play a crucial role in the performance of the whole system. As to sellers, the procedure is quite simple since they sell products required and disappear when the stock gets exhausted. As to buyers, instead, the algorithm is rather more complex. At each turn they must interrogate another buyer, buy something from a seller, and possibly answer a question from another buyer. Each of these actions leads to a number of decisions to be taken.

- Buying. Here the question to be answered is which seller a buyer should turn to. The easiest option would be to pick the seller with the best image, or (in L2) the best reputation if image is not available. A threshold is set for an evaluation (actually, for its center of mass, see [**?**] for definitions) to be considered good enough and be used for choosing. In addition, a limited chance to explore other sellers is possible, as controlled by the system parameter risk 3. Notice that image has always priority over reputation, since unlike reputation image implies that the evaluation is shared by the user.

– Asking. As in the previous case, the first choice to be made is which agent to be queried, and the decision making procedure is exactly the same as that for choosing a seller, but now agents deal with images and reputations of targets as informers (informer image) rather than sellers. Once decided whom to ask, the question is what to ask. Only two queries are allowed:
  - Q1 - Ask information about a buyer as informer (basically, how honest is buyer X as informer), and
  - Q2 - Ask for some good or bad seller (for instance, who is a good seller, or who is a bad seller). Notice that this second question does not refer to one specific individual, but to the whole body of information that the queried agent may have. This is in order to allow for managing large numbers of sellers, when the probability to choose a target seller that the queried agent has information about would be low. The agent will ask one of these two questions with a probability of 50%. If Q1 is chosen, buyer X as informer would be the least known, i.e., one with less information to build up an image or reputation about.
– Answering. Let agent S be the agent asking the question, R the agent being queried. Agents can lie, either because they are cheaters or because they are retaliating. When a buyer is a cheater, they provide information after having turned its value into the opposite. Retaliation is accomplished by sending inaccurate information (for instance, sending *I-dont-know* when it has information, or simply giving the opposite value) when R has got a bad image of S as informer. In L1 retaliation is done by sending a *I-dont-know* message even when R has got information. This avoids possible retaliation from S since a *I-dont-know* message implies no commitment. If reputation is allowed, (L2) retaliation is accomplished in the same way as if the agent were a liar, except that image is converted into reputation in order to avoid potential retaliations from S. Fear of retaliation leads to sending an image only when agent is certain about evaluation. This is yet another parameter (Strength) allowing the fear of retaliation to be implemented. Notice that if strength is null, there is no fear since any image will be a candidate answer, no matter what its strength is. As strength increases, agents become more conservative, with less image and more reputation circulating in the system.

6.3 Expected Results

Based on the hypthesis that image allows for more retaliation than reputation, we expect the following results to obtain:

– H1 Initial advantage: L2 shows an initial advantage over L1, that is, L2 grows faster.
– H2 Performance: L2 performs better as a whole, that is, the average quality at regime is higher than L1.

Some questions concerning cheating and fairness were also investigated:

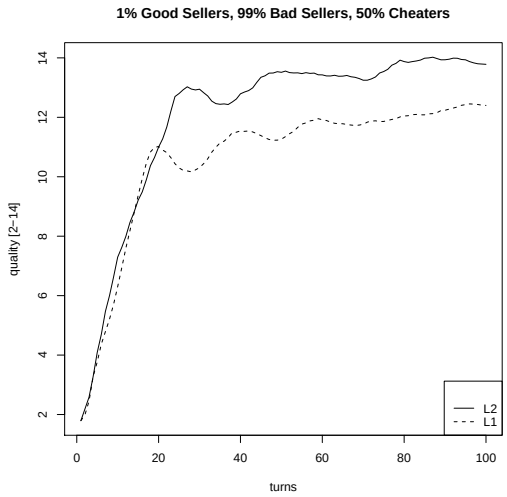

Fig. 1. Accumulated average quality per turn in condition A1 (very few good sellers), 50% informational cheaters, for L1 and L2 agents. L2 agents show better performance even with large amount of false information.

- cheaters' advantage: do cheaters effectively reach a significant advantage thanks to their behavior?
- Cheaters' effects: are cheaters always detrimental to the system? In particular, is the performance of the system always decreasing as a function of the number of cheaters?

Simulations to enquire on the relationship between L1 and L2 has been run with the following parameters: with fixed stock (50), number of buyers (25), and number of sellers (100); different values of informational cheaters (percentages of 0%, 25% and 50%); different values of bad sellers, ranging from the extreme case of 1% of good vs 99% of bad sellers (A1), going trough 5% good sellers Vs 95% bad sellers (A2), 10% good sellers vs 90% bad sellers (A3), and finally, to another extreme where we have 50% of good sellers vs 50% of bad sellers (A4). Note that from A1 to A4 the maximum level of quality obtainable increases (from experimental data, we move from a regime maximum quality of about 14 in A1 to nearly full quality in A4). For each one of these conditions and for every situation (L1 and L2) we run 10 simulations. In the figures we present the accumulated average earnings per turn in both situations, L1 and L2. In L1 the amount of useful communications (different from *I-dont-know*) is much lower that in L2, due to the fear of retaliation that governs this situation. In conditions where communication is not important, the difference between the levels disappears.

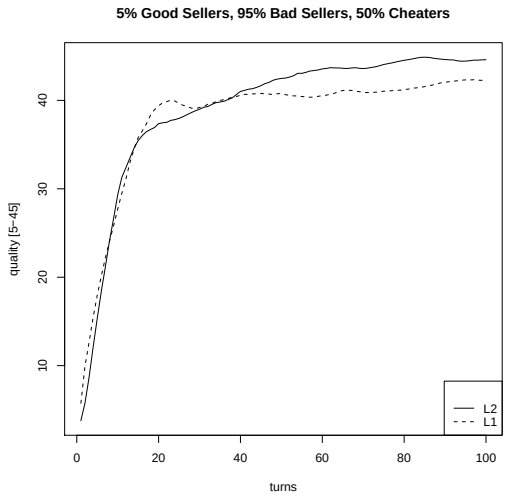

Fig. 2. Accumulated average quality per turn in condition A2 (5% good sellers), 50% informational cheaters, for L1 and L2 agents. The margin of L2 on L1 is reduced.

In the following, we report only the result of the experiment with cheaters, where the difference between L1 and L2 is made more significative by the presence of false information. For a full report, please refer to [**?**].

6.4 Experiments with Cheaters

We report results of experiments with 50% of informational cheaters in conditions A1, A2, A3 and A4. The large amount of false information produces a bigger impact in situations and conditions where communication is more important. Quality reached in L1 shows almost no decrease with respect to the experiment without cheaters, while L2 quality tends to drop to L1 levels. This shows how the better performance of L2 over L1 is due to the larger amount of information that circulates in L2. In Figure **??**, nothwithstanding the large amount of false information, there is still a marked difference between the two levels. Essentially, L2 agents show a better performance in locating the very rare good sellers. The situation starts to change in Figure **??**, where the two algorithms are more or less comparable; here, the larger amount of good sellers does not make necessary the subleties of L2. In Figure **??**, with an even larger amount of good sellers available, the two algorithms show the same level of performance.

7 Conclusions and Future Work

Results indicate that reputation plus image (as opposed to image only) improves the average quality of products exchanged in the whole system. The value added of reputation is shown under the occurrence of

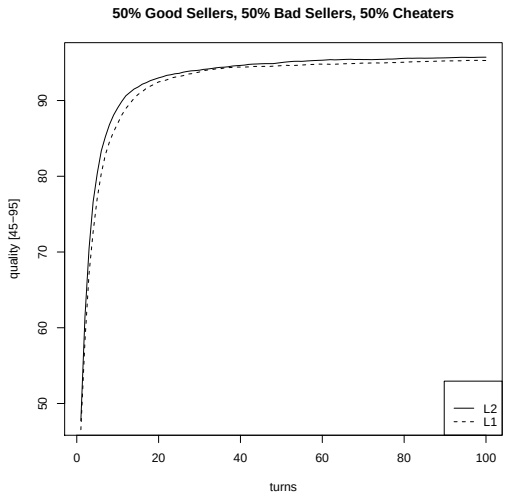

Fig. 3. Accumulated average quality per turn in condition A4 (half good sellers), 50% informational cheaters, for L1 and L2 agents. The two levels are indistinguishable.

- retaliation: personal commitment associated to image transmission exposes the agent to possible retaliation if inaccurate information was sent. Conversely, reputation transmission does not lead to such a consequence, but at the same time provides agents with information that might be useful to select satisfactory partenrs. Future work will concern the effect of cheaters over the whole system in presence of a norm that prescribes agents to tell the truth. The reputation mechanism will turn into a social control artifact aimed to identify and isolate agents that do not follow that norm.
- Communication: There is no reputation without communication. Therefore, scenarios with no or poor communication are irrelevant for the study of reputation. However, in virtual societies with autonomous communicating agents that need to cooperate and are enabled to choose partners, reputation considerably increases the circulation of information and improves the performance of their activities. In our experiments, even when there is no penalty for direct interactions and only one question per turn is allowed, the introduction of reputation improves the average quality per turn. In scenarios where quality is scarce and agents are completely autonomous this mechanism of social control makes the difference.
- Decision making procedure: The decision making model implemented has a decisive impact on the system's performance. In fact, this is where the agent may take advantage of the distinction between image and reputation. In future work, we will elaborate on this distinction, possibly reformulating it in terms of textitmeta decision making, a very promising future line of work to better ground and exploit the image and reputation artefacts.

These results gives us reasons to draw some more general conclusions about the respective role of image and reputation. The antisocial consequence of image spreading seems to be clearly documented in the experiment we have reported upon. But if this is the case, we also find evidence for our argument that social networks based upon image perform more poorly than networks based upon reputation at least when partner selection is a common goal of the network members. An image-network, based upon acquaintanship, if not familiarity, and trusted communication of own evaluations, stimulates retaliation or at least discrimination when informers are found to spread incorrect information. Consequently, such a type of network shows poor robustness against not only deception and cheating, but also errors and rumour.

Conversely, reputation-based networks are more flexible and inclusive, they tolerate errors. Though selecting information before using it, the reputation mechanism does not lead recipients to discard so easily nor, a fortiori, retaliate against bad informers. In such a way, the chain of retaliations is prevented and the consequent lowering of the exchanges' quality is reduced. These considerations apply for our simplified, structureless network. It would be interesting, in future works, to examine also the effects of network connectivity, applying for example the "sever tie unilaterally, create tie consensually strategy employed in [**?**].

Finally, does such a view of reputation point to an account of the evolution of socially desirable behaviour, concurrent with the classical one, based on punishment and strong reciprocity (cf. [**?**])? Hard to say for the time being. However, this is a fascinating resarch hypothesis for future studies.

Acknowledgements

We thank Isaac Pinyol for simulation deployment and execution. We recognize the help and encouragement of the reviewers, whose comments have stimulated us to improve this paper.

References

1. Marmo, S.: L'uso della reputazione nelle applicazioni internet: prudenza o cortesia? l'approccio socio-cognitivo. In: AISC - Terzo Convegno Nazionale di Scienze Cognitive. (2006)
2. Tadelis, S.: What's in a name? reputation as a tradeable asset. The American Economic Review (1999)
3. Conte, R., Paolucci, M.: Reputation in artificial societies: Social beliefs for social order. Kluwer Academic Publishers (2002)
4. Pinyol, I., Paolucci, M., Sabater-Mir, J., Conte, R.: Beyond accuracy. reputation for partner selection with lies and retaliation. In: MABS 07, Eighth International Workshop on Multi-Agent-Based Simulation. (2007)
5. Sabater, J., Paolucci, M., Conte, R.: Repage: Reputation and image among limited autonomous partners. Journal of Artificial Societies and Social Simulation **9**(2) (2006)

6. Conte, R., Castelfranchi, C.: Cognitive Social Action. London: UCL Press (1995)
7. Conte, R.: Social intelligence among autonomous agents. Computational and Mathematical Organization Theory **5** (1999) 202–228
8. Leslie, A.M.: Pretense, autism, and the 'theory of mind' module. Current Directions in Psychological Science **1** (1992) 18–21
9. Miceli, M., Castelfranchi, C. In: The Role of Evaluation in Cognition and Social Interaction. Amsterdam:Benjamins (2000)
10. Castelfranchi, C., Poggi, I.: Bugie, finzioni e sotterfugi. Per una scienza dell'inganno. Carocci Editore, Roma. (1998)
11. Sen, S., Sajja, N.: Robustness of reputation-based trust: boolean case. In: AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM Press (2002) 288–293
12. Izquierdo, S.S., Izquierdo, L.R.: The impact of quality uncertainty without asymmetric information on market efficiency. Journal of Business Research **60**(8) (August 2007) 858–867
13. Hanaki, N., Peterhansl, A., Dodds, P.S., Watts, D.J.: Cooperation in evolving social networks. Management Science **(in press)**
14. Fehr, E., Fischbacher, U., Gächter, S.: Strong reciprocity, human cooperation and the enforcement of social norms. Human Nature **13** (2002) 1–25