Chapter 15
Reputation

Francesca Giardini, Rosaria Conte, and Mario Paolucci

Why Read This Chapter? To understand the different conceptions underlying reputation in simulations up to the current time and to get to know some of the approaches to implementing reputation mechanisms, which are more cognitively sophisticated.

Abstract In this chapter, the role of reputation as a distributed instrument for social order is addressed. A short review of the state of the art will show the role of reputation in promoting (a) social control in cooperative contexts – like social groups and subgroups – and (b) partner selection in competitive contexts, like (e-) markets and industrial districts. In the initial section, current mechanisms of reputation – be they applied to electronic markets or MAS – will be shown to have poor theoretical backgrounds, missing almost completely the cognitive and social properties of the phenomenon under study. In the rest of the chapter a social cognitive model of reputation developed in the last decade by some of the authors will be presented. Its simulation-based applications to the theoretical study of norm-abiding behaviour, partner selection and to the refinement and improvement of current reputation mechanisms will be discussed. Final remarks and ideas for future research will conclude the chapter.

15.1 Reputation in Social Systems: A General Introduction

Ever since hominid settlements started to grow, human societies needed to cope with the problem of social order (Axelrod, 1984): how to avoid fraud and cheating in wider, unfamiliar groups? How to choose trustworthy partners when the likelihood of re-encounter is low? How to isolate cheaters and establish worthwhile alliances with the good guys?

F. Giardini (✉) • R. Conte • M. Paolucci
Laboratory for Agent-Based Social Simulation (LABSS), Institute of Cognitive Sciences and Technologies, National Research Council, Via Palestro, 32 – 00185, Rome
e-mail: francesca.giardini@istc.cnr.it; rosaria.conte@istc.cnr.it; mario.paolucci@istc.cnr.it

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_15,

Social knowledge like reputation and its transmission, i.e. gossip, plays a fundamental role in social order, adding at the same time cohesiveness to social groups and allowing for distributed social control and sanctioning (plus a number of other functionalities, see Boehm 1999). Reputation is a property that unwilling and unaware individuals derive from the generation, transmission and manipulation of a special type of social beliefs, namely social evaluations, and that contributes to regulate natural societies from the morning of mankind (Dunbar 1996). People use reputational information to make decisions about possible interactions, to evaluate candidate partners, to understand and predict their behaviours, and so on.

That reputation is a fundamental generator, vehicle and manipulator of social knowledge for enforcing reciprocity and other social norms is known since long (see a review in Conte and Paolucci 2002). In particular, in the study of cooperation and social dilemmas, the role of reputation as a partner selection mechanism started to be appreciated in the early 1980s (Kreps and Wilson 1982). However, little understanding of its dynamic and cognitive underpinnings was achieved at that stage. Despite its critical role in the enforcement of altruism, cooperation and social exchange, the social cognitive study of reputation is relatively new. Hence, it has not yet been fully clarified how this critical type of knowledge is manipulated in the minds of agents, how social structures and infrastructures generate, transmit and transform it, and consequently how it affects agents' behaviour.

The aim of this chapter is to guide the reader through the multiplicity of computational approaches concerned with the impact of reputation and its dynamics. Reputation is a complex social phenomenon that cannot be treated as a static attribute of agenthood, with no regard for the underlying process of transmission. We claim that reputation is both the process and the effect of transmitting information, and that further specifications about the process and its mechanisms are needed. Following these premises, we will first review some applications of reputation in computational simulation, highlighting problems and open questions, and then we will propose a theoretical social cognitive model of reputation. Moreover, we will present three different ways of applying the cognitive theory of reputation to model social phenomena: the Sim-Norm model, the SOCRATE framework and the REPAGE architecture.

This brief introduction will be followed by an outline of reputation research in different domains (social psychology, management and experimental economics, agent-based simulation), in order to show how many different viewpoints can be used to describe and explore this complex phenomenon. We will then focus on some of the results in electronic markets and multi-agent simulations. Electronic markets are a typical example of a complex environment where centralized control is not possible and decentralized solutions are far from being effective. In recent years, the Internet contributed to a growth of auction sites facilitating the exchange of goods between individual consumers, without guaranteeing transparency and safety of transactions. On the other hand, multi-agent applications are concerned with the problem of assessing the reliability of single agents and of social networks. In Sect. 15.2 we will propose a cognitive model of reputation, which aims to solve

some of the problems left open by existing systems, moving from a theoretical analysis of cognitive underpinnings of reputation formation and spreading. This model will be tested in the following section, where a description of three different implementations will be provided. Finally, we will draw some conclusions about future work and directions.

15.1.1 State of the Art

According to Frith and Frith (2006), there are three ways to learn about other people: through direct experience, through observation and through "cultural information". When the first two modalities are not available, reputational information becomes essential in order to obtain some knowledge about one's potential partner(s) in an interaction, and thus to predict their behaviour. Reputation allows people to predict, at least partially or approximately, what kind of social interaction they can expect and how that interaction may possibly evolve. Reputation is therefore a coordination device whose predictive power is essential in social interactions (Paolucci and Conte 2009). Furthermore, reputation has a strategic value and can be used to pursuit self-interest (Paine 1967; Noon and Delbridge 1993).

Reputation and its transmission (gossip) has an extraordinary preventive power: it substitutes personal experience in (a) identifying cheaters and isolating them, and in (b) easily finding trustful partners. It makes available all the benefits of evaluating someone, without implying the costs of direct interaction.

Furthermore, in human societies gossip facilitates the formation of groups (Gluckman 1963): gossipers share and transmit relevant social information about group members within the group (Barkow 1996), at the same time isolating out-groups. Besides, gossip contributes to stratification and social control, since it works as a tool for sanctioning deviant behaviours and for promoting, even through learning, those behaviours that are functional with respect to the group's goals and objectives. Reputation is also considered as a means for sustaining and promoting the diffusion of norms and norm conformity (Wilson et al. 2000).

The theories of indirect reciprocity and costly signals show how cooperation in large groups can emerge when the agents are endowed with, or can build, a reputation (Nowak and Sigmund 1998a; 1998b; Gintis et al. 2001). As Alexander (1987) pointed out, "indirect reciprocity involves reputation and status, and results in everyone in the group continually being assessed and reassessed". According to this theory large-scale human cooperation can be explained by individuals helping others in order to uphold a reputation and thus be included in future cooperation (Panchanathan and Boyd 2004).

Despite important advances in the study of reputation as a means to support cooperation (Sommerfeld et al. 2008), no explicit theory of the cognitive ingredients and processes which reputation is made of was provided. More recently,

reputation and gossip has started to become crucial in other fields of the social sciences like management and organisation science, governance, or business ethics, where the importance of branding became apparent. The economic interest in the subject matter implied an extension of reputation to the super-individual level: Corporate reputation is considered as an external and intangible asset tied to the history of a firm and coming from stakeholders' and consumers' perceptions. Rose and Thomsen (2004) claim that good reputation and financial performance are mutually dependent, hence a good reputation may influence the financial asset of a firm and vice versa. Several researchers have tried to create a corporate reputation index containing the most relevant dimensions to take into account when dealing with corporate reputation. Cravens et al. (2003) interviewed 650 CEO in order to create a reliable index, but their index has so many entries, ranging from global strategy to employees' attributes, that it is not easy to foresee how such a tool could be used. Gray and Balmer (1998) distinguish between corporate image and corporate reputation. Corporate image is the mental picture consumers hold about a firm, therefore is similar to an individual perception, whereas the reputation results from the firm's communication and long-term strategy. Generally speaking, corporate reputation is treated as an aggregate evaluation stakeholders, consumers, managers, employees, and institutions form about a firm, but the mechanisms leading to the final result are still only vaguely defined.

Over the last 10 years several studies in experimental economics have investigated reputational dynamics through standard experimental settings, such as trust games, public good games, and moral hazard problems (see Fehr and Gachter 2000, for an introduction). The aim of these studies is to explain mechanisms underlying reciprocity, altruism and cooperation in humans, in order to answer the puzzling question: "Why do humans cooperate?" Axelrod (1984). According to Nowak and Sigmund (1998a, b), reputation sustains the emergence of indirect reciprocity, which gets people to cooperate in order to receive cooperation even from strangers. Following this hypothesis, holding a good reputation in your social group makes it more probable that someone else would help you when you will need help.

If social order is a constant of human evolution, it becomes particularly crucial in the e-society where the boundaries of interaction are extensively widened. The portentous development pace of ICT technologies dramatically enlarges the range of interaction among users, generating new types of aggregation, from civic communities to electronic markets, from professional networking to e-citizenship, etc. What is the effect of this widening of social boundaries? Communication and interaction technologies modify the range, structures and modalities of interaction, with consequences that are only partially explored, often only to resume the stereotype of technological unfriendliness (see the negative impact of computer terminals, as opposed to face-to-face interaction, on subjects' cooperativeness in experimental studies of collective and social dilemmas, Sell and Wilson 1991; Rocco and Warglien 1995). A detailed approach to the effects of technological infrastructures on interaction styles and modes has never been adopted. Perhaps, an exception to

this is represented by the research on the effects of asymmetry of information on the markets. Known to encourage fraud and low-quality production, this phenomenon appears as an intrinsic feature of e-markets, but in fact it goes back to eleventh century Maghribi traders moving along the coast of the Mediterranean sea (Greif 1993). As exemplified by Akerlof (1970), asymmetry of information drives honest traders and high quality goods out of the market. The result is a market where only "lemons", or fraudulent commodities, are available – often to the detriment of both sellers and buyers. The classical example of a lemon market is the used car market, where only sellers have information about problems with the cars they are selling, and most consumers are incapable of discerning these problems. Contemporary online traders such as users of Internet auction sites face the same problem: online buyers can learn about the quality (or condition) of the good only once they have already paid for it.

Auction sites may be very generic concerning the products being offered and operate on a global scale (e.g., e-Bay), or may focus on specific products on a national scale (many car auction sites). At the moment virtually all consumer products are being auctioned on the Internet, ranging from used toys and CDs to cars and houses Utz et al. (2009). Compared to buying through online retailers, buying through auction sites is even less controllable, as the sellers are not visible and have not made major investments. Consumers who purchase through auction sites must rely on the accuracy and reliability of the seller (Melnik and Alm 2002). Sellers on the Internet may actively try to communicate their reputation to potential buyers, inflating common expectations on the effects of reputation. Melnik and Alm (2002) investigated whether an e-seller's reputation matters. Their results indicated that reputation had a positive – albeit relatively small – impact on the price levels consumers were willing to pay. Moreover, Yamagishi et al. (2009) show that reputation has significant positive effects on the quality of products.

Despite the role of reputation in economic transactions, online reputation reporting systems are only moderately efficient (Resnick and Zeckhauser 2002; Bolton et al. 2002). eBay, one of the best (if not the best) known examples, is also one of the world's largest online marketplaces with a community of over 50 million users registered. On eBay, most items are sold through English-type auctions, and the reputation mechanism used is based on the ratings that users perform after the completion of transactions. The user can give three possible values: positive (1), negative (-1) or neutral (0). The "reputation" value is the sum of the last 6 months' ratings, weighted by the relevance of the transaction.

In all of these models, the notion of reputation is weak and essentially reduced to centralized image: no direct exchange of information takes place among participants but only reports to a central authority, which calculates the final reputation value. The actual utility of this mechanism is debatable. For example, when forums are available, real reputation exchanges are performed in parallel, ignoring the centrally calculated reputation rating. Moreover, people are likely to not provide reputational feedback (under-provision) and if they do, they lean on providing only good reports (over-scoring). In sum, the reputation system does not seem to perform efficiently.

However, from an economic point of view, eBay prospers. How is this possible? What would happen if reputation worked more efficiently? Which type of reputation system should be used for which objective? As current online reputation systems are not theory-driven instruments, based upon an understanding of what reputation is, how it is generated and distributed, and how it works, all of these questions are still open.

However, the effects of reputation are at least partially known, generally restricted to partner selection. Indeed, agent-based social simulation has taught us some lessons: (1) what matters about reputation is its transmission (Castelfranchi et al. 1998), since by this means agents acquire zero-cost relevant information; (2) reputation has more impact than image: if agents transmitted only their own evaluations about one another (image), the circulation of social knowledge would stop soon (Pinyol et al. 2008). To exchange information about reputation, agents need to participate in circulating reputation whether they believe it or not (gossip) and, to preserve their autonomy, they must decide how, when and about whom to gossip. What is missing in the study of reputation is the merging of these separate directions in an interdisciplinary integrated approach, which accounts for both its social cognitive mechanisms and structures.

15.1.2 Simulating Reputation: Current Systems

So far, the simulation-based study of reputation has been undertaken for the sake of social theory, namely in the account of pro-social behaviour – be it cooperative, altruistic, or norm-abiding – among autonomous, i.e. self-interested agents.

In this chapter, we will concentrate instead on two specific functionalities of reputation:

1. To promote norm-abiding behaviour in cooperative settings (e.g. social groups).
2. To favour partner selection in electronic markets and agentized environments.

The second aspect has been dealt with in the literature to some extent and the remainder of this section will give an overview of existing approaches; for a treatment of the first aspect we refer the reader to Sect. 15.2, where we mainly discuss our own models and studies of reputation.

Several attempts have been made to model and use reputation in artificial societies, especially in two sub-fields of information technologies: computerized interaction (with a special reference to electronic marketplaces) and agent-mediated interaction. It is worth emphasizing that in these domains trust and reputation are actually treated as the same phenomenon, and often the fundamentals of reputation mechanisms are derived from trust algorithms. Moreover, several authors (Moukas et al. 1999; Zacharia 1999; Zacharia et al. 1999) explain reputation in terms of trust and vice versa, continuously mixing up these two phenomena. We will review some

of the main contributions in online reputation reporting systems and in multi-agent systems, in order to achieve a better understanding of the complex issue of implementing and effectively using reputation in artificial societies.

15.1.2.1 Online Reputation Reporting Systems

The continuously growing volume of transactions on the World Wide Web and the growing number of frauds this appears to entail[1] led scholars from different disciplines to develop new online reputation reporting systems. These systems are meant to provide a reliable way to deal with reputation scores or feedbacks, allowing agents to find cooperative partners and avoid cheaters.

The existing systems can be roughly divided into two sub-sets: Agent-oriented individual approaches and agent-oriented social approaches, depending on how agents acquire reputational information about other agents.

The *agent-oriented individual approach* has been dominated by Marsh's ideas on trust (Marsh 1992, 1994a, b), on which many further developments and algorithms are based. This kind of approach is characterized by two attributes: (1) any agent may seek potential cooperation partners, and (2) the agent only relies on its experiences from earlier transactions. When a potential partner proposes a transaction, the recipient calculates the "situational reputation" by weighing the reputation of his potential trading partner with further factors, such as potential output and the importance of the transaction. If the resulting value is higher than a certain "cooperation threshold", the transaction takes place and the agent updates the reputation value according to the outcomes of the transaction. If the threshold is not reached, the agent rejects the transaction offer, which may be punished by a "reputation decline". These individual-based models (Bachmann 1998; Marsh 1994a; Ripperger 1998) differ with regard to the length of memory span they apply. Agents may forget their experiences slowly, fast, or never.

In *agent-oriented social approaches* agents not only rely on their direct experience, but are also allowed to consider third-party information (Abdul-Rahman and Hailes 1997a; Rasmusson 1996; Rasmusson and Janson 1996; Schillo 1999; Yu and Singh 2000). Although these approaches share the same basic idea, i.e. experiences of other agents in the network can be used when searching for the right transaction partner, they rely upon different solutions when it comes to weigh the third-party information and to deal with "friends of friends". Thus the question arises: how to react to information from agents who do not seem to be very trustworthy?

Another problem lies in the storage and distribution of information. To form a complete picture of its potential trading partners, each agent needs both direct (its own) and indirect (third-party) evaluations in order to be able to estimate the validity and the informational content of such a picture.

[1] The US-based Internet Crime Complaint Center (IC3) received 231,493 complaints for the year 2005, 62.7 % of which were related to electronic auctioning (IC3 2005).

Regan and Cohen (2005) propose a system for computing indirect and direct reputation in a computer mediated market. Buyers rely on reputation information about sellers when choosing from whom to buy a product. If they do not have direct experience from previous transactions with a particular seller they take indirect reputation into account by asking other buyers for their evaluations of the potential sellers. The received information is then combined to mitigate effects of deception. The objective of this system is to propose a mechanism which reduces the "undesirable practices" on actual reputation in online applications, especially on the part of sellers, and to prevent the market from turning into a "lemons market" where only low quality goods are listed for sale.

One serious problem with this and similar models concerns the reputation transmission. Agents only react to reputation requests, while proactive, spontaneous delivery of reputation information to selected recipients is not considered. On the other hand, despite its simplicity, this type of model tackles the problem of collusion between rating agents by keeping secret the evaluation of sellers amongst buyers, i.e. not disclosing it to the sellers.

As to electronic marketplaces, classic systems like eBay show a characteristic bias to positive evaluations (Resnick and Zeckhauser 2002), suggesting that factual cooperation among users at the information level may lead to a "courtesy" equilibrium (Conte and Paolucci 2003). As Cabral and Hortaçsu (2006) formally prove, initial negative feedbacks trigger a decline in sale price that drives the targeted sellers out of the market. Good sellers, however, can gain from 'buying a reputation' by building up a record of favourable feedback through purchases rather than sales. Thus those who suffer a bad reputation stay out – at least until they decide to change identity – while those who stay in can take advantage of a good reputation: after a good start, they will hardly receive negative feedback and even if they do, negative feedbacks will not get to the point of spoiling their good name. Under such conditions, even good sellers may have an incentive to sell lemons.

Intuitively, the courtesy equilibrium reduces the deterrent effect of reputation. If a reputation system is meant to impede frauds and improve the quality of products, it needs to be constructed in such a way as to avoid the emergence of a courtesy equilibrium. It is not by chance that among the possible remedies to ameliorate eBay, Dellarocas (2003) suggested a short-memory system, erasing all feedbacks but the very last one. While this might work for eBay, we believe that such remedies based on fragmented models and tailored to a particular application are not the way forward. Instead, a general theory of how reputation and its transmission work needs to be developed. On top of such a theory, different systems for different objectives can then be constructed. We will pursue this further in Section 15.3.

15.1.2.2 MAS Applications

Models of trust and reputation for multi agent systems applications (e.g. Yu and Singh 2000; Carbo et al. 2002; Sabater and Sierra 2002; Schillo et al. 2000; Huynh et al. 2004; for exhaustive reviews see Ramchurn et al. 2004a; Sabater and Sierra 2005)

present interesting new ideas and advances over conventional online reputation systems, with their notion of centralized global reputation.

Yu and Singh (2000) proposed an agent-oriented model for social reputation and trust management which focuses on electronic societies and MAS. Their model introduces a gossip mechanism for informing neighbours of defective transaction partners, in which the gossip is transferred incrementally through the network of agents. It also provides a mechanism to include other agents' testimonies in an agent's reputation calculation. Agents store information about the outcome of every transaction they had and recall this information in case they are planning to bargain with the same agent again (direct evaluation). If the agent meets an agent it has not traded with before, the reputation mechanism comes into play. In this mechanism, so-called referral chains are generated that can make third-party information available across several intermediate stations. An agent is thus able to gain reputation information with the help of other agents in the network. Since a referral chain represents only a small part of the whole network, the information delivered will most likely be partial instead of global as in centralized systems like eBay.

In the context of several extensive experiments, Yu and Singh showed that the implementation of their mechanism results in a stable system, in which the reputation of cheaters decreases rapidly while the cooperating agents experienced a slow, almost linear increase in reputation. Still, some problems remain. The model does not combine direct and indirect reputation, i.e. if an agent has already traded with another agent he has to rely on his own experience and cannot use the network information anymore. Thus it might take unnecessarily long to react to a suddenly defecting agent that cooperated before. In addition, Singh and Yu do not give an explanation of how their agent-centred storage of social knowledge (for example the referral chains) is supposed to be organized. Consequently, no analysis of network load and storage capacity can be done.

As this example shows, the "agentized environment" is likely to produce interesting solutions that may apply also to online communities. This is so for two main reasons. First, in this environment two problems of order arise: to meet the users' expectations (external efficiency), and to control agents' performance (internal efficiency). Internal efficiency is instrumental to the external one, but it re-proposes the problem of social control at the level of the agent system. In order to promote the former, agents must control, evaluate, and act upon each other. Reliability of agents is a proxy for reliability of users. Secondly, and consequently, the agent system plays a double role: it is both a tool and a simulator. In it one can perceive the consequences of given premises, which may be transferred to the level of user interactions. In a sense, implemented agent systems for agent-mediated interaction represent both parallel and nested sub-communities.

As a consequence, solutions applied to the problems encountered in this environment are validated more strictly, against both external and internal criteria. Their effects are observable at the level of the virtual community, with a procedure essentially equivalent to agent-based simulation and with the related advantages. Moreover, solutions may not (only) be implemented between agents, but (also) within agents, which greatly expands the space for modelling. So far, however,

these potentials have not been fully exploited. Models have mainly been aimed at ameliorating existing tools implemented for computerized markets. We suggest that agent systems can do much more than this: they can be applied to answer the question as to (a) what type of agent, (b) what type of beliefs, and (c) what type of processes among agents are required to achieve useful social control. More specifically, what type of agents and processes are needed for which desirable result: better efficiency, encouraging equity and hence users' trust, discouraging either positive or negative discrimination (or both), foster collaboration at the information level or at the object level (or both), etc.

15.1.3 Concluding Remarks

This review has given an overview of how reputation have been discussed and modelled in studies regarding online markets and multi-agent systems.

In the Internet-based models, the notion of reputation is weak and essentially reduced to centralized image: participants do not exchange information directly but only report their evaluations to a central authority, which calculates the final reputation value. The actual utility of this mechanism is debatable.

The solutions proposed for MAS systems are interesting but so far insufficient to meet the problems left open by online systems. There is a tendency to consider reputation as an external attribute of agents without taking into account the process of creation and transmission of that reputation.

We argue that a more theory-driven approach is needed, based upon a conceptual analysis of the differences and analogies among notions concerning social evaluation. In the next section we will therefore introduce our social cognitive approach towards reputation.

15.2 An Alternative Approach: The Social Cognitive Theory of Reputation

In this section we will present a social cognitive model of reputation, we will define the difference between image and reputation, introduce the roles different agents play when evaluating someone and transmitting this evaluation and, finally, we will explain the decision processes underlying reputation.

Let us first clarify the term "social cognitive". 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 an in-depth discussion of these notions, see Conte and Castelfranchi 1995; Conte 1999).

Social cognitive processes are 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. Leslie 1992), which focuses upon one, albeit important aspect of social agency, namely social beliefs, this approach aims at modelling and possibly implementing systems acting in a social – be it natural or artificial – environment. Thus, it 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).

15.2.1 Image and Reputation

The social cognitive model presented here 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. We will call this evaluation the social *image* of the agent. An agent's *reputation* is argued to be distinct from, although strictly interrelated with, its image. Image consists of a set of evaluative beliefs (Miceli and Castelfranchi 2000) about the characteristics of the target, i.e. it is an assessment of her positive or negative qualities with regard to a norm, a competence, etc. Reputation is both the process and the effect of transmission of a target image. The image relevant for social reputation may concern a subset of the target's characteristics, e.g. its willingness to comply with socially accepted norms and customs. More precisely, we define reputation to consist of three distinct but interrelated objects: (1) a cognitive representation, i.e. a believed evaluation; (2) a population object, i.e. a propagating believed evaluation; (3) an objective emergent property at the agent level, i.e. what the agent is believed to be.

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 (Conte and Paolucci 2002). In particular, it proceeds from the level of individual cognition to the level of social propagation (population level) and from there back to individual cognition. What is even more interesting, once it reaches the population level it gives rise to an additional property at the agent level. From the very moment an agent is targeted by the community, their life will change whether they want or believe it or not. Reputation has become the immaterial, more powerful equivalent of a scarlet letter sewn to one's clothes. It is more powerful because it may not even be perceived by the individual to whom it is attached, and therefore it is not in the individual's power to control and manipulate. Reputation is an objective social property that

emerges from a propagating cognitive representation. This lack of an identified source, i.e. its impersonality, is the distinctive feature of reputation, whereas image always requires identifying the individual who made the evaluation.

To formalise these concepts we will begin by defining the building block of image. An agent has made an evaluation when he or she believes that a given entity is good for, or can achieve a given goal. An agent has made 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:

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

Often, the sets of evaluators and beneficiaries overlap 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 B, which may or may not include the evaluator. Evaluations may concern physical, mental, and social properties of targets; agents may evaluate a target with regard to both capacity and willingness to achieve a shared goal. The latter, willingness to achieve a goal or interest, is particular to social evaluations. Formally, e (with $e \in E$) may evaluate t ($t \in T$) with regard to a state of the world that is in b's ($b \in B$) 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 of B, or it may favour a supra-individual entity, that results from the interactions among the members of B (for example, if B's members form a team).

To make this analysis more concrete, we will start with an example drawn from the Sim-Norm model, which will be described in more detail in Section 15.3.1. Let us consider a classical multi-agent situation, a set of agents fighting for access to a scarce resource (food). Assume that a norm of precedence (a proscription against attacking agents who are consuming their "own" resources) is applied to reduce conflicts. The norm is disadvantageous for the norm follower in the short run, but is advantageous for the community, and thus eventually for the individual followers. We will call N the set of norm followers, or normative agents, and C the set of cheaters, or violators of the norm. With regard to social evaluations (image), the targets coincide with the set of all agents; $T = N \cup C$ (all are evaluated). The agents carrying out the evaluation are restricted to the norm followers: $E = N$, because only the normative find purpose in evaluating. Finally, $B = N \cup C$: indeed, if normative agents benefit globally from the presence of the norm, cheaters in this simple setting benefit even more by exploiting the norm; they can attack the weaker while they themselves are safe from attacks by the gullible normative.

It is very easy to find examples where all three sets coincide. General behaviour norms, such as "Do not commit murder" apply to, benefit, and are evaluated by the whole universe of agents. There are situations in which beneficiaries, targets, and evaluators are separate, for example, when norms safeguard the interests of a subset of the population. Consider the quality of TV programs for children broadcast in the

afternoon. Here, we can identify three more or less distinct sets. Children are the beneficiaries, while adults entrusted with taking care of children are the evaluators. It could be argued that *B* and *E* still overlap, since *E* may be said to adopt *B*'s interests. The targets of evaluation are the writers of programs and the decision-makers at the broadcast stations. There may be a non-empty intersection between *E* and *T* but no full overlap. In case the target of evaluation is the broadcaster itself, a supra-individual entity, the intersection can be considered to be empty: $\mathrm{E} \cap \mathrm{T} = \emptyset$.

Extending this formalisation to include reputation, we have to differentiate further. 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 therefore involves four sets of agents:

1. A nonempty set *E* of agents who share the evaluation;
2. A nonempty set T of evaluation targets;
3. A nonempty set *B* of beneficiaries, i.e. the agents sharing the goal with regard to which the elements of *T* are evaluated;
4. 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 underlying belief or not.

15.2.2 Reputational Roles

Agents may play more than one role simultaneously: evaluator, beneficiary, target, and third party. In the following, we will examine the characteristics of the four roles in more detail.

15.2.2.1 Evaluator

Any autonomous agent is a potential evaluator Conte et al. 1998. Social agents are likely to form evaluative beliefs about one another (see Castelfranchi 1998) as an effect of interaction and social perception. These may be positive or negative, depending on an agent's experiences. When agents evaluate one another with regard to their individual goals they obtain social evaluations. Image, the result of such social evaluations, serves to identify friends and partners and to avoid enemies. Who should the agent resort to for help? Who should he or she cooperate with? And who should be avoided due to being dangerous or ill willed?

Furthermore, agents may not only evaluate one another with regard to their own goals, but with regard to the goals or interests of a given set of agents (the beneficiaries), to which the evaluators may belong. A negative evaluation may be formed about agents violating others' rights or behaving in an apparently malevolent and hostile manner, whether or not the evaluators consider themselves potential victims of such actions. Information thus obtained may be used to infer that the target could violate other rights in the future, namely, those of the evaluator. In addition, evaluators may be concerned with one another's power to achieve the goals or interests of abstract social entities or institutions, as when we judge others' attitudes towards norms, the church, the government or a political party.

To sum up, agents evaluate one another with regard to their own goals and the goals they adopt from either other individual agents (e.g. their children) or supra-individual agents, such as groups, organisations, or abstract social entities.

15.2.2.2 Beneficiary

A beneficiary is the entity that benefits from the action with regard to which targets are evaluated. Beneficiaries can either be individual agents, groups and organisations or even abstract social entities like social values and institutions. Beneficiaries may be aware of their goals and interests, and of the evaluations, but this is not necessarily the case. In principle, their goals might simply be adopted by the evaluators – as it happens, for example, when members of the majority support norms protecting minorities. Evaluators often are a subset of the beneficiaries.

Beneficiaries may be implicit in the evaluation. This is particularly the case when it refers to a social value (honesty, altruism, etc.); the benefit itself and those who take advantage of it are left implicit, and may coincide with the whole society. The beneficiary of the behaviour under evaluation is also a beneficiary of this evaluation: the more an (accurate) evaluation spreads, the likelier the execution of the positively evaluated behaviour.

15.2.2.3 Target

The target of any social evaluation is the evaluated entity. While targets may even be objects or artefacts to be used by others when the evaluation pertains to image, for reputation mental and moral components are necessarily involved. Holders of reputation (targets) are endowed with the following important characteristics:

- Agency, in particular autonomous agency and sociality (the target is evaluated with regard to a given behaviour)
- Mental states, specifically willingness to perform the respective behaviour
- Decision making or deliberative capacity, i.e. the ability to choose a desirable behaviour from a set of options
- Social responsibility, i.e. the power to prevent social harm and possibly to respond for it, in case any damage occurs.

Intuitively, targets are complementary to beneficiaries, but this is not necessarily the case. Targets may be evaluated for their willingness to exhibit a behaviour from which they are expected to benefit themselves.

Other than beneficiaries, targets are always explicit. They may be individual entities or supra-individual like a group, a collective, an abstract entity, or a social artefact, such as an institution, provided this can be attributed the capacity to make decisions, achieve goals, and perform actions. A further distinguishing characteristic of targets is that they may inherit the reputation of the set they belong to. Offspring may inherit their parents' reputation; employees may suffer from the bad reputation of the firm they work for; members may inherit the reputation of their social groups, whether they have had a chance to decide to join those groups or not. In the latter case, the attitude under evaluation is considered to be a built-in tendency shared by the targets.

Two dynamic processes characterise reputation. The first concerns the propagation of a social cognitive representation from one agent to another. We call this *transmission of reputation* (or gossip). As we have seen so far, it is intrinsic to reputation spreading. The second process concerns the extension of a given agent's reputation to other agents who are related to the former by affiliation, proximity, similarity, etc. We will call it *contagion of reputation* (or inheritance). This dynamic is not intrinsic to reputation, although it may empirically co-occur with reputation spreading. It corresponds to what is usually called prejudice.

The transmission process spreads reputation as a cognitive representation, whereas the contagious process spreads reputation as a property. Both may proceed vertically and horizontally, within social groups and hierarchies and from one agent to his or her neighbours in social space. In this sense, reputation may be perceived as a highly fertile phenomenon, whose propagation is favoured by social membership and proximity.

15.2.2.4 Third Party or Gossiper

An agent is a (potential) third party if she transmits (is in position to transmit) reputation information about a target to another agent or set of agents. Although sharing awareness of a given target reputation, third parties do not necessarily share the corresponding image (social evaluation) of the target. That is, they do not necessarily believe it to be true. Actually, an agent could share a given evaluation with others without being aware of others' existence. This would put him in a poor position to transmit reputation. If there is no belief about E, it is very difficult to justify (that is, to assign a goal to) the act of reputation transmission. Instead, the meta-belief about the sharing of the evaluation is a powerful tool and paves the way to complex (and potentially deceiving) new behaviours.

Third parties (if they are also targets) may deserve a negative evaluation; they may actually deceive their fellow third parties, the beneficiaries, the targets, or the society as a whole, by conveying information that they hold to be false. A third party may be bluffing: he or she may pretend to be benevolent with regard to

beneficiaries, in order to (1) enjoy the advantages of sharing reputation information, (2) be considered as part of the in-group by other evaluators, and therefore (3) gain a good reputation without sustaining the costs of its acquisition (as would be implied by performing the socially desirable behaviour), and (4) avoid the consequences of a bad reputation. Agents may also spread a false reputation, i.e., pretend that a target has a given reputation when this is not the case. Agents do this in order to achieve the aforementioned benefits without taking responsibility for spreading a given social evaluation.

Agents may have incomplete and inaccurate information – a situation that can result in ill-reputed agents being taken as well-reputed and vice versa. One reason for incomplete or inaccurate information is the lack of personal experience and familiarity with the target. In such a case, an agent might have received information from other third parties or from agents who have had a direct contact with the target.

To demonstrate how the social categories so far identified may help in the analysis of reputation spreading we will return to the example of the quality of TV programs for children. Transmission of evaluations involves two distinct sets of agents, children (beneficiaries) and adults with the children's welfare in mind (evaluators), while the targets – operators and decision-makers at broadcast stations – are a subset of the evaluators. The set of third parties comprises the whole universe of adults, thereby including evaluators and targets but not the beneficiaries. The targets (broadcaster) offer an interesting example of the interplay between institutional and personal roles. While no assumption of benevolence towards *B* can be made on the part of the institution (broadcast stations) itself, whose purpose is profit, things are more intriguing in the case of their employees. These belong to *E* and are thus benevolent with regard to B, but are also members of *T* and thus committed to their employer.

15.2.3 Reputation-Based Decisions

After identifying the different roles agents can take with regard to image, reputation and its transmission let us now examine the relevant decision-making processes. To understand the difference between image and reputation, we will categorise the mental decisions based upon them into three levels:

1. The *epistemic level* is concerned with decisions about accepting or rejecting an evaluative belief.
2. The *pragmatic–strategic level* concerns decisions about interacting with another agent (target).
3. The *memetic level* concerns decisions about transmitting evaluative beliefs about a given target to others

15.2.3.1 Epistemic Level

When an agent decides whether or not to accept a particular belief that forms either a given image or acknowledges a given reputation he makes an epistemic decision. This involves direct evaluation of the belief in question. To acknowledge a given reputation implies two more specific beliefs: (a) that a nonempty set *E* of agents within the same population shares the given image about the target (reputation effect), and/or (b) that information about the target's image has been circulated (reputation process) among those agents.

Image and reputation about one and the same agent do not necessarily overlap. From an agent *X*'s meta-belief that agent *Y* is said to be e.g. a womaniser we are not entitled to derive *X*'s belief that *Y* is in fact a womaniser (although *X* may draw such a more or less arbitrary conclusion). Image and reputation about the same *Y* may be consistent (for example, *X* believes that *Y* is and is believed to be a womaniser) or inconsistent (*X* believes that *Y* either suffers from an undeserved bad reputation or enjoys an unworthy good one).

To accept one does not imply acceptance of the other. The two processes of acceptance are different not only in their respective outputs (evaluations and meta-beliefs), but also in the operations involved. To accept a given image implies coming to share it. The acceptance may be based, for example, upon supporting evidence and first-hand experience with the image target, consistent pre-existing evaluations (concerning, for example, the class of objects to which the target belongs), or trust in the source of the given evaluative belief.

Acknowledging a given reputation, on the other hand, does not necessarily lead to sharing others' evaluations but rather to believing that these evaluations are held or at least 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 for him or her to hear some rumours.

15.2.3.2 Pragmatic-Strategic Level

Agents resort to their evaluative beliefs in order to achieve their goals (Miceli and Castelfranchi 2000). In general, evaluations are guidelines for planning; therefore social evaluations (evaluations about other agents) are guidelines for social action and social planning. The image an agent *X* has about a target agent *Y* will guide his or her action with regard to *Y*, will suggest whether it is convenient to interact with *Y* or not, and also will suggest what type of interaction to establish with *Y*. Image may also be conveyed to others in order to guide their actions towards the target in a positive or negative sense. To do so, the agent must (pretend to) be committed to their evaluation and take responsibility for its truth value.

While image dominates direct pragmatic-strategic decisions reputation may be used when it is consistent with image or when no image of the target has been

formed. To influence others' social decisions, however, agents tend to transmit information about the target's reputation rather than their image of the target. Two main reasons explain this inverse pattern:

Agents expect that a general opinion 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 have to take responsibility for it; consequently, they may influence others at a lower personal cost.

15.2.3.3 Memetic Level

A memetic decision can be roughly described as the decision to spread reputation. Consequently, communication about reputation is 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. In other words, communication about reputation does neither imply any personal commitment of the speaker with regard to the content of the information delivered – if an agent *X* reports on *Y*'s bad reputation he is by no means stating that *Y* deserved it – nor any responsibility with regard to the credibility of (the source of) information ("I was told that *Y* is a bad guy"). This is due to the source of the meta-belief being implicit ("I was told. . .") and the set of agents to whom the belief is attributed being undefined ("*Y* is ill/well reputed").

Communication about reputation is not always sincere. On the contrary, the lack of commitment and responsibility can (and often does) lead to deceptive reputation transmission. To deceive other agents about *Y*'s reputation agent *X* only needs to report it as a rumour independent of or even opposite to his own beliefs.

Several factors may affect an agent's decision to transmit reputation information (see Conte and Paolucci 2002):

- Certainty and acceptance of the evaluation,
- Reputation of the source from which information was received,
- Responsibility and accountability for the effects of distributing this evaluation to others,
- Benevolence toward the beneficiary as opposed to benevolence toward the target, or no benevolence at all.

Social responsibility can be defined as the power attributed to intelligent autonomous social systems to predict and prevent harm to others and/or themselves. A memetic agent may be said to have harmed another agent by transmitting bad reputation information about her. Responsibility comes from accountability, i.e. the power and obligation to respond and repair harm that one has been found accountable for. A memetic agent may be asked to take responsibility and possibly repair the

harm. However, responsibility is much less crucial in a memetic than in a pragmatic decision concerning reputation. The decision to transmit reputation will thus depend on the extent to which the memetic agent is aware of potentially serious effects of reputation transmission on the target and his perceived contribution to these effects. The latter is influenced by the perceived "distance" from the target. It is easier to convey potentially dangerous evaluations when the target is far away, absent, unknown, etc., not only because the memetic agent is less likely to perceive the effective occurrence of harm but also because he or she will not be a direct cause of it. Harm will actually depend on a number of intermediate events, others' decisions, etc. The greater the distance and the smaller the memetic agent's (perceived) responsibility, the less cautious the memetic agent is likely to be.

Benevolence means whether and to what extent a candidate third party adopts the goals or interests of either the target or the beneficiary of reputation transmission. When circulating reputation, memetic agents may follow different strategies, according to the direction of their benevolence.

In case of benevolence towards the set of beneficiaries B, gossiping may follow some prudence rule like "pass on negative evaluation even if uncertain, pass on positive evaluation only if certain". This may give way to circulation of a reputation which is overly critical and worse than the real characteristics of the target.

When the benevolence is target-oriented (set T), it is possible to expect the application of some courtesy rule like "pass on positive evaluation even if uncertain, negative evaluation only if certain". This may lead to a courtesy equilibrium where no-one expresses critique anymore, especially fearing retaliation.

If memetic agents are not benevolent towards any of the two sets B and T, Conte and Paolucci (2002) predicts scarcity of reputation transmission. In this case production of information may be induced by institutional reinforcement of it, e.g. a university requesting students to academically evaluate their lecturers.

Systematic application of a courtesy or prudence rule in reputation spreading may result in selective transmission of either positive or negative evaluations. We expect general adoption of such rules as a consequence of self-interested decisions of single memetic agents. We may reasonably suppose that a lower responsibility of memetic agents increments the quantity of circulating information, while benevolence direction is a consequence – all other factors being equal – of the overlapping of reputational roles of agents. Role overlapping means that the same people (or people having the same norm-related goals) are involved in more than one reputational role. Such agents then may consider it useful to be prudent or generous in their information spreading, or they may have not enough motivation to circulate information of any sort.

As we have seen in this and the previous section, operationalization of cognitive properties of reputation and dynamics of reputational groups allows us to express testable hypotheses, which can be investigated both by experimenting with human subjects and by designing software reputation agents.

15.3 Simulating Reputation

Social simulation models have been successfully employed to investigate the effects of reputation in different contexts. In this section we will present an overview of three different ways of designing agent-based systems based on the cognitive theory of reputation presented previously. These models can be distinguished according to:

- Different levels of cognitive complexity of the agents,
- The kind of setting (purely cooperative, competitive or both)
- The scenario.

The first, Sim-Norm, based on a very simple concept of reputation, has been applied to observe the impact of reputation in social control. The second, REPAGE, based on a more complex agent architecture and on the concepts of reputation and image introduced in the previous section, has been applied to explore the impact of social evaluation on the market. Finally, in the third model, called SOCRATE, relatively simple agents interact in a complex market in which they exchange both goods and information.

15.3.1 Sim-Norm

This model was developed to examine the effect of reputation on the efficiency of a norm of precedence (Castelfranchi, Conte, Paolucci 1998; Conte and Paolucci 1999; Paolucci 2000) in reducing aggression, measured both at the global (i.e. societal) and local (i.e. individual) level. In particular, Sim-Norm was designed to explore why self-interested agents exercise social control. Albeit far from reaching a final conclusion on this issue, the studies based on Sim-Norm confirmed a positive impact of reputation on social control.

More precisely, while individually acquired evaluation of other agents gave norm executors no significant advantage, the transmission of these evaluations among norm executors proved decisive in levelling the outcomes of norm-abiders and cheaters (if numerically balanced).

15.3.1.1 Hypotheses

Sim-Norm revolved around the question of which ingredients are necessary for social order to be established in a society of agents. The role of norms as aggression controllers in artificial populations living under conditions of resource scarcity was addressed. We set out to explore two hypotheses:

Norm-based social order can be maintained and its costs reduced via distributed social control.

Social cognitive mechanisms are needed to account for distributed social control. In particular, the propagation of social beliefs plays a decisive role in distributing social control at low or zero individual costs and high global benefit.

15.3.1.2 Model Description and Experimental Conditions

The model defines agents as objects moving in a two-dimensional environment (a 10 × 10 grid) with randomly scattered food. At the beginning of a run, agents and food items are assigned locations at random. A location is a cell in the grid. The same cell cannot contain more than one object at a time (except when an agent is eating). The agents move through the grid in search of food, stopping to eat to build up their strength when they find it. The agents can be attacked only when eating; no other type of aggression is allowed.

At the beginning of each step, every agent selects an action from the six available routines: *eat*, *move-to-food-seen*, *move-to-food-smelled*, *attack*, *move-random*, and *pause*. Actions are supposed to be simultaneous and time consuming.

To investigate the role of norms in the control of aggression, we compared scenarios in which agents follow a norm – implemented as a restriction on attacks – with identical scenarios, in which they follow utilitarian rules. In all scenarios, each agent can perform only one of three strategies:

Blind aggression, or control condition, in which aggression is not constrained. If the agent can perform no better move (eating, moving to food seen or smelled), then it will attack without further considerations. Blind agents have access to neither their own strength nor the eater's strength; these parameters never enter their decision-making process.

Utilitarian, in which aggression is constrained by strategic reasoning. Agents will only attack those eaters whose strength is lower than their own. An eater's strength is "visible" one step away from the agent's current location. While blind agents observe no rule at all, utilitarian agents observe a rule of personal utility, which does not qualify as a norm.

Normative (N), in which aggression is constrained by a norm. We introduced a finder-keeper precept, assigning a "moral right" to food items to finders, who become possessors of the food. Possession of food is ascribed to an agent on the grounds of spatial vicinity; food owned is flagged, and every player knows to whom it belongs. Each food unit may have up to five owners, decided on the basis of proximity at the time of creation. The norm then prescribes that agents cannot attack other agents who are eating their own food.

The strategies can also be characterised by the kind of agents they allow to attack: while blind agents attack anybody, the utilitarian agents attack only the weaker, and the normative agents, respecting a norm of private property, will not attack agents who are eating their own food.

These strategies were compared (Castelfranchi et al. 1998) using an efficiency measure – the average strength of the population after *n* periods of simulation – and a fairness measure, the individual deviation from the average strength.

15.3.1.3 Findings

The first two series of experiments showed that normative agents perform less well than non-normative agents in mixed populations, as they alone bear the costs of social control and are exploited by utilitarian agents. In a following series of experiments, "image" was added to the preceding experimental picture. Now, each normative agent collects information about the behaviour of other agents in an image vector. This information is binary and discriminates between the "respectful," who will abide with the norm and "cheaters," who will not respect the principle of finders–keepers. The vector is initialised to "all respectful" (presumption of innocence), but every time a normative agent is attacked while eating its own food, the attacker is recorded as a cheater. Moreover, the normative algorithm is modified so that the agents respect the norm only when facing agents known as respectful, while they behave with known cheaters according to one of the retaliation strategies listed above.

The results from another set of experimental runs on a mixed population equally composed of normative and utilitarian agents show that in this model useful knowledge can be drawn from personal experience, but therefore still at one's own cost. To reduce cost differences among subpopulations, image is insufficient.

Henceforth, we provided the respectful agents with the capacity to exchange with their (believed-to-be) respectful neighbours at distance one images of other agents. With the implementation of a mechanism of transmission of information, we can speak of a reputation system. We ran the experiments again with normative agents exchanging information about cheaters. The results suggest that circulating knowledge about others' behaviours significantly improves normative agents' outcomes in a mixed population.

The spreading of reputation can then be interpreted as a mechanism of cost redistribution for the normative population. Communication allows compliant agents to easily acquire preventive information, sparing them the costs of direct confrontations with cheaters. By spreading the news that some guys cheat, the good guys (1) protect themselves, (2) at the same time punish the cheaters, and possibly (3) exercise an indirect influence on the bad guys to obey the norm. Social control is therefore explained as an indirect effect of a "reciprocal altruism" of knowledge.

15.3.1.4 Concluding Remarks

The Sim-Norm model presented in this section was studied with the purpose of clarifying the role of norms controlling aggression in simple multi-agent systems. The model shows that a simple norm of precedence is not as efficient as a utilitarian

rule when utilitarian agents are present; the normative agents must resort to retaliation against cheaters. The addition of image is not enough to defend the norm, but image coupled with a mechanism of information transmission is. The necessity of information transmission points out the relevance of our distinction between image and reputation.

The model inspired further research in the social simulation community: Saam and Harrer (1999) used the same model to explore the interaction between normative control and power, whereas Hales (2002) applied an extended version of SimNorm to investigate the effects of group reputation. In his model agents are given the cognitive capacity to categorise other agents as members of a group and project reputation onto whole groups instead of individual agents (a form of stereotyping).

15.3.2 The REPAGE Cognitive Architecture

REPAGE is a computational system for partner selection in a competitive setting (marketplace), although in principle it can also be used in cooperative contexts (organizations). Based on a model of REPutation, imAGE and their interplay, REPAGE provides evaluations of potential partners and is fed with information from others plus outcomes from direct experience. This is fundamental to account for (and to design) limited autonomous agents as exchange partners. To select good partners, agents need to form and update own social evaluations; hence, they must exchange evaluations with one another.

But in order to preserve their autonomy, agents need to *decide* whether or not to share others' evaluations of a given target. If agents would automatically accept reported evaluations and transmit them as their own, they would not be autonomous anymore. In addition, in order to exchange information about reputation, agents need to participate in circulating it, whether they believe it or not (gossip); but again to preserve their autonomy, they must *decide* how, when and about whom to gossip.

In sum, the distinction between image and reputation suggests a way out from the paradox of sociality, i.e. the trade-off between agents' autonomy and their need to adapt to social environment. On one hand, agents are autonomous if they select partners based on their social evaluations (images). On the other, they need to update evaluations by taking into account others' evaluations. Hence, social evaluations must circulate and be represented as "reported evaluations" (reputation), before and in order for agents to decide whether to accept them or not. To represent this level of cognitive detail in artificial agents' design, there is a need for a specialised subsystem. This is what REPAGE provides.

In the following we briefly describe the architecture of the REPAGE model and how it is integrated with the other elements that compose a typical deliberative agent (cf. Sabater et al. 2006 for a more detailed description). REPAGE is composed of three main elements: a memory, a set of components called detectors and the analyzer. An implementation of REPAGE in Java has been published as a Sourceforge project.

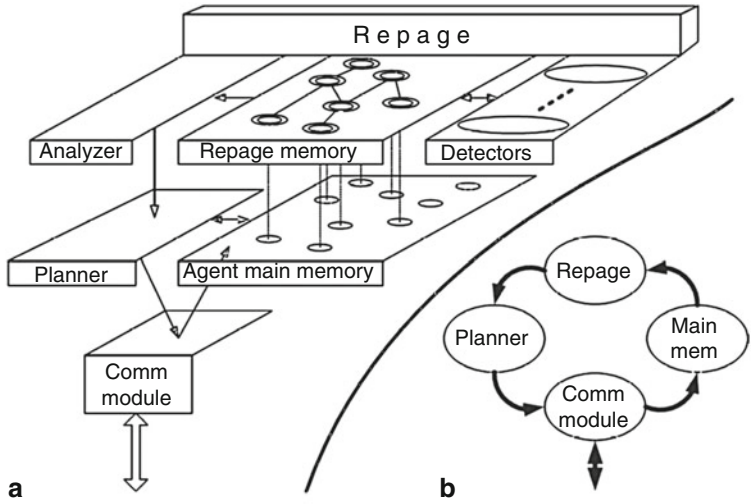

Fig. 15.1 REPAGE and its environment. For the sake of clarity only agent components that interact with REPAGE are depicted

15.3.2.1 Memory

In the implementation, to support the specialized nature of REPAGE, memory is actually composed by a set of references to the predicates in the agent general-purpose memory (see section A of Fig. 15.1). Only those predicates that are relevant for dealing with image and reputation are considered. Therefore, of all the predicates in the main memory, only a subset is also part of the REPAGE memory. A change in a predicate is immediately visible in both memories.

To mirror their dependence connections, in the REPAGE memory predicates are conceptually organized in different levels and inter-connected. Each predicate reference is wrapped by a component that adds connection capabilities to it.

Predicates contain a fuzzy evaluation, consisting of three aspects: the type of the evaluation (e.g. personal experience, image, third party image), the role of the target (e.g. informant or seller) and the actual content. To store the content, a simple number is used in e-Bay and in most reputation systems. This sharp representation, however, is quite implausible in inter-agent communication, which is one of the central aspects of REPAGE; in real life no-one is told that people are saying Jane is 0.234 good. To capture the lack of precision coming (a) from vague utterances, e.g. "I believe that agent *X* is good, I mean, very good – good, that is", and (b) from noise in the communication or in the recollection from memory, we decided to model the actual value of an evaluation with a tuple of positive real values that sum to one. Aggregation of these evaluations was detailed in Sabater and Paolucci (2007).

Finally, each predicate has a strength value associated to it. This value is a function of (1) the strength of its antecedents and of (2) some special characteristics intrinsic to that type of predicate. The network of dependencies specifies which

predicates contribute to the values of other predicates. Each predicate in the REPAGE memory has a set of antecedents and a set of consequents. If an antecedent changes its value or is removed, the predicate is notified. Then the predicate recalculates its value and notifies the change to its consequents.

15.3.2.2 Detectors

The detectors are inference units specialized in certain predicates. They populate the REPAGE memory (and consequently the main memory of the agent) with new predicates inferred from those already in the memory. They are also responsible for removing predicates that are no longer useful and, more importantly, for creating the network of dependencies among the predicates.

Each time a new predicate is added to or removed from the main memory (either by the action of another agent module – planner, communication module, etc. – or by the action of a detector) the REPAGE memory notifies all detectors 'interested' in that type of predicate. This starts a cascading process where several detectors are activated one after the other. At the same time, the dependency network ensures the correct update of the predicate values according to the new additions and subtractions.

15.3.2.3 Analyzer

The main task of the analyzer is to propose actions that (1) can improve the accuracy of the predicates in the REPAGE memory and (2) can solve cognitive dissonances trying to produce a situation of certainty. The analyzer can propose one or more suggestions to the planner, which then decides whether to execute them or not.

15.3.2.4 Integration with Deliberative Agent Architecture

One of the key points of the REPAGE design is its easy integration with the other elements that compose a deliberative agent. REPAGE is not only a passive module the agent can query to obtain information about image and reputation of another agent. The aim of REPAGE is also to provide the agent (or more specifically, the planner module of the agent) with a set of possibilities that can be followed to improve the reliability of the provided information.

The communication module connects the agent with the rest of the world. After a possible process of filtering and/or transformation of the received information, new predicates are added to the main memory.

The REPAGE memory contains references to those predicates in the main memory of the agent that are relevant to deal with image and reputation. The actions of the detectors over the REPAGE memory result in addition/removal of

predicates as well as the creation of the dependence network. While the addition or removal of predicates has again an immediate effect on the main memory, the dependence network is present only in the REPAGE memory.

The planner uses the information in the main memory to produce plans. This information includes the information generated by REPAGE. By means of the analyzer, REPAGE always suggests new actions to the planner in order to improve the accuracy of existing images and reputations. It is a task of the planner to decide which actions are worth being performed. These actions (usually asking informers or interacting with other agents) will hopefully provide new information that will feed REPAGE and improve its accuracy. This cycle is illustrated in Fig. 15.1(B).

15.3.2.5 Demonstration

To illustrate some of the main points in the behaviour of REPAGE let us consider two particular situations that are quite common in the application area of markets. The general scenario is the following: Agent *X* is a buyer who knows that agent *Y* sells what he needs but knows nothing about the quality of agent *Y* (the target of the evaluations) as a seller. Therefore, he turns to other agents in search for information – the kind of behaviour that can be found, for example, in Internet fora, auctions, and in most agent systems.

In the first situation, agent *X* receives a communication from agent *Z* saying that his image of agent *Y* as a seller is very good. Since agent *X* does not yet have an image about agent *Z* as informer he resorts to a default image that is usually quite low. The uncertain image as an informer adds uncertainty to the value of the communication.

Later on, agent *X* has received six communications from different agents containing their image of agent *Z* as an informer. Three of them give a good report and three a bad one. This information is enough for agent *X* now to build an image about agent *Z* as an informer so this new image substitutes the default candidate image that was used so far. However, the newly formed image is insufficient to take any strategic decision – the target seems to show an irregular behaviour.

At this point, agent *X* decides to try a direct interaction with agent *Y*. Because he is not sure about agent *Y* he resorts to a low risk interaction. The result of this interaction is completely satisfactory and has important effects in the REPAGE memory. The candidate image about agent *Y* as a seller becomes a full image, in this case a positive one.

Moreover, this positive image is compared (via a fuzzy metric) with the information provided by agent *Z* (which was a positive evaluation of agent *Y* as a seller); since the comparison shows that the evaluations are similar, a positive confirmation of the image of agent *Z* as an informer is generated. This reinforcement of the image of agent *Z* as a good informer at the same time reinforces the image of agent *Y* as a good seller. As a consequence, there is a positive feedback between the image of agent *Y* as a good seller and the image of agent *Z* as a good informer. This feedback is a necessary and relevant part of the REPAGE model.

The purpose of the second situation is to show how REPAGE differentiates between image and reputation. In this case agent *X*, after a couple of successful interactions with agent *Y*, receives four communications from different informants. Each informant communicates the reputation of agent *Y* as a seller, which happens to be negative. This contradicts agent *X*'s own positive evaluations of Y stemming from their direct interactions. However, it is not a problem in REPAGE because there is a clear distinction between image and reputation. In addition, unlike communicated images (see first situation) communications about reputation do not generate confirmations that reinforce or weaken the image of the informant.

15.3.2.6 Concluding Remarks

REPAGE is a cognitive architecture to be used by artificial autonomous agents in different scenarios, be they MAS applications, artificial markets or teamwork. Its main objective is to give agents access to a fundamental module for social reasoning and decision-making on the first two levels described in Sect. 15.2.3.1 and 15.2.3.2, epistemic and strategic. Thus, the main outputs of REPAGE consist of "advice" about what to think of a given target and how to interact with it.

However, there are still some aspects of reputation-based decisions, which need further theoretical elaboration and inclusion into REPAGE. First, the interplay between image and reputation in the epistemic decision: which received evaluations will be transformed into own evaluations of a given target and why? Secondly, the grounds on which memetic decisions are made: what is said to whom, and moreover, why? Both these issues will prove particularly relevant in testing this architecture by means of simulation.

15.3.3 SOCRATE

SOCRATE is an attempt to test the cognitive theory of reputation in a ideal-typical economic setting, modeled after an industrial district in which firms exchange goods and information (Giardini et al. 2008; Di Tosto et al. 2010). In this model, the focus is on social relationships among agents and their impact on the focusing on social links and on the resulting social structure, usually informal, which is a defining feature of industrial clusters (Porter 1998; Fioretti 2005; Squazzoni and Boero 2002). Social evaluations are the building blocks of social and economic relationships inside the cluster; they are used to select trustworthy partners, to create and enlarge the social network (Giardini and Cecconi 2010), and to exert social control on cheaters.

15.3.3.1 Agents and Environment

We implemented an artificial environment in which agents can choose among several potential suppliers by relying either on their own evaluations, or on other agents' evaluations. In the latter case, the availability of truthful information could help agents to find reliable partners without bearing the costs of potentially negative, i.e. harmful, interactions with bad suppliers. Moreover, evaluations can be transmitted either as image (with an explicit source and the consequent risk of retaliation) or as reputation.

We tried to answer the following questions:

- How relevant is image when firms need to select suppliers, service providers and so on?
- Does transmission of image promote the improvement of quality in a cluster?
- How does false information affect the quality of the cluster?
- What are the effects of image and reputation, respectively, on the economic performance of firms?

Our model is characterized by the existence of two different kinds of interactions among agents: material exchange and evaluation exchange. The former refers to the exchange of products between leader firms and their suppliers, and it leads to the creation of a supply chain network. On the other hand, the flows of social evaluations among the firms create a social network. In this setting, agents can transmit true or false evaluations in order to either help or hamper their fellows searching for a good partner.

Agents are firms organized into different layers, in line with their role in the production cycle. The number of layers can vary according to the characteristics of the cluster, but a minimum of two layers is required. Here, we have three layers, but n possible layers can be added, in order to develop a more complex production process:

Layer 0 (L0) is represented by leader firms that supply the final product
Layer 1 (L1) is represented by suppliers of L0
Layer 2 (L2) are firms providing raw material to firms in L1.

When image transmission is allowed, both leader firms and suppliers exchange information with their fellows, thus creating and taking part in a social network. This process works only horizontally: L0 and L1 are not allowed to talk each other. Agents in both layers can play two possible roles:

Questioner – asks an Informer, i.e. another firm of the same layer, to suggest a good supplier;

Informer – provides her own image of a good supplier. Honest informers suggest their best rated supplier, whereas cheaters transmit the image of their worse supplier (as if it was a good one).

Agents in L0 have to select suppliers that produce with a quality above the average among all L1 agents. Suppliers can be directly tested or they can be chosen thanks to the information received by other L0 firms acting as Informers. Buying products from L1 and asking for information to L0 fellows are competing activities that can not be performed contemporaneously. In turn, once received an order for a product, L1 firms should select a good supplier (above the average quality) among those in L2. After each interaction with a supplier, both L0 and L1 agents create an evaluation, i.e. an image, of it, comparing the quality of the product they bought with the quality threshold value. Agents are endowed with an "Image Table" in which all the values of the tested partners are recorded and stored for future selections. In the Reputation condition, evaluations are exchanged without revealing their source, thus injecting the cluster with untested information. In this condition, retaliation against untrustful informers is unattainable.

At each simulation cycle, firms attempt to interact with the best known suppliers. Every time the best known supplier is unavailable they query their fellows about other high quality suppliers, which will be tested and integrated into the Image-Table.

15.3.3.2 Results

The exchange of true and valuable information led to a significant increase in quality of exchanged goods, making the cluster exploration faster and permitting firms to obtain higher profits. However, in the Image condition average quality was negatively affected by the presence of cheaters, because false information triggered a mechanism of reciprocal retaliation with detrimental effects on the cluster as a whole. In the Reputation Condition, the cluster could absorb relatively high percentages of cheaters, without compromising its economic performance.

SOCRATE results provided further support to the hypotheses about the importance of reputation for social control, showing again that social evaluations and their features have consequences also in economic terms.

15.3.3.3 Concluding Remarks

Given the assumption that in this "small-world", as in the real world, evaluations are crucial to selecting trustworthy partners and to isolating cheaters, we tried to demonstrate how useful this exchange is, especially in terms of global cluster quality and profits. Firms receiving reliable information about potential partners found good suppliers in a faster and more efficient way, compared to firms that were systematically cheated by their fellows. More interesting results are expected after we include an enriched economic structure and the implementation of reputation.

15.4 Conclusion and Future Work

In the last decade there has been a significant increase in research on reputation and gossip. There is growing evidence on the fact that the presence of reputation strongly promotes cooperation and represents an effective way to maintain social control. Exercising social control roughly means to isolate and punish the cheaters. However, punishment is costly and it inevitably implies the problem of second-order cooperation.

In this chapter, we discussed current studies of reputation as a distributed instrument for social order. After a critical review of current technologies of reputation in electronic institutions and agentized environments, a theory of reputation as a social cognitive artefact was presented. In this view, reputation allows agents to cooperate at a social meta-level, exchanging information (a) for partner selection in competitive settings like markets and (b) for cheater isolation and punishment in cooperative settings like teamwork and grouping.

To exemplify both functionalities, we introduced three simulation models of reputation in artificial societies developed within our research group during the last decade. Both have been used mainly as a theory-building tool.

The first, Sim-Norm, is a reputation-based model for norm compliance. The main findings from simulations show that, if circulated among norm-abiders only, reputation allows for the costs of compliance to be redistributed between two balanced subpopulations of norm-abiders and cheaters. In such a way, it contributes to the fitness of the former, neutralising the advantage of cheaters. However, results also show that as soon as the latter start to bluff and optimistic errors begin to spread in the population, things worsen for norm-abiders, to the point that the advantage produced by reputation is nullified.

REPAGE, a much more complex computational model than SimNorm, was developed to test the impact of image, reputation and their interaction on the market. Based on our social cognitive theory, it allows the distinction between image and reputation to be made, and the trade-off between agents' autonomy and their liability to social influence to be coped with. REPAGE allows the circulation of reputation whether or not third parties accept it as true.

Finally, SOCRATE is an attempt to combine fairly complex agents (endowed with a memory and able to manage different kinds of evaluations) with a market in which agents must protect themselves from both informational and material cheating. In this context, reputation has been proven to be useful to punish cheaters but it also prevented the social network from collapse.

These results clearly show that differentiating image from reputation provides a means for coping with informational cheating and that further work is needed to achieve a better understanding of this complex phenomenon. The long term results of these studies are expected to (a) answer the question on how to cope with informational cheating (by testing the above hypothesis), (b) provide guidelines about how to realize technologies of reputation that achieve specified objectives (e.g. promoting

respect of contracts vs. increasing volume of transactions), and finally (c) show the impact of reputation on the competitiveness of firms within and between districts.

Acknowledgements The authors would like to thank Jordi Sabater and Samuele Marmo for their helpful collaboration. This work was partially supported by the Italian Ministry of University and Scientific Research under the Firb programme (SOCRATE project, contract number RBNE03Y338), by the European Community under the FP6 programme (eRep project, contract number CIT5-028575; EMIL project, contract number IST-FP6-33841).

Further Reading

For a more in-depth treatment of the contents of this chapter we refer the reader to the monograph *Reputation in Artificial Societies* (Conte and Paolucci 2002). A review of this book was published in JASSS (Squazzoni 2004). For more on the same line of research, with an easier presentation aimed to dissemination, we suggest the booklet published as the result of the eRep project (Paolucci et al. 2009).

A conference aiming to propose a scientific approach to Reputation has been organized in 2009: the first International Conference on Reputation, ICORE 2009. Its proceedings (Paolucci 2009), available online, contain a collection of papers that give an idea of the range of approaches and ideas on Reputation from several academic disciplines.

Due to the focus on the theoretical background of reputation only a narrow selection of simulation models of reputation could be discussed in this chapter. Sabater and Sierra (2005) give a detailed and well-informed overview of current models of trust and reputation using a variety of mechanisms. Another good starting point for the reader interested in different models and mechanisms is the review by Ramchurn and colleagues (Ramchurn et al. 2004a).

Further advanced issues for specialised reputation subfields can be found in (Jøsang et al. 2007), a review of online trust and reputation systems, and in (Koenig et al. 2008), regarding the Internet of Services approach to Grid Computing.

References

Abdul-Rahman A, Hailes S (1997a) A distributed trust model. In: Proceedings of NSPW97, new security paradigms workshop VI, Langdale, Cumbria, 23–26 Sept 1997. ACM Press, New York, pp 48–60

Akerlof G (1970) The market for lemons: quality uncertainty and the market mechanisms. Q J Econ 84:488–500

Alexander R (1987) The biology of moral systems. De Gruyter, New York

Axelrod R (1984) The evolution of cooperation. Basic Books, New York

Bachmann R (1998) Kooperation, vertrauen und macht in systemen verteilter künstlicher intelligenz: eine vorstudie zum verhältnis von soziologischer theorie und technischer modellierung. In: Malsch T (ed) Sozionik. Edition Sigma, Berlin, pp 197–234

Barkow JH (1996) Beneath new culture is old psychology: gossip and social stratification. In: Barkow JH, Cosmides L, Tooby J (eds) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, New York, pp 627–638

Bohem C (1999) Hierarchy in the forest: the evolution of egalitarian behavior. Harvard University Press, Cambridge, MA

Bolton G, Katok E, Ockenfels A (2002) How effective are online reputation mechanisms? An experimental investigation (Working paper). Max Planck Institute of Economics, Jena

Cabral L, Hortaçsu A (2006) The dynamics of seller reputation: theory and evidence from eBay (Discussion papers, 4345). Centre for Economic Policy Research, London

Carbo J, Molina JM, Davila J (2002) Comparing predictions of SPORAS vs. a fuzzy reputation agent system. In: Grmela A, Mastorakis N (eds) Proceedings of the 3rd WSEAS international conference on Fuzzy Sets and Fuzzy Systems (FSFS '02), Interlaken, 11–15 Feb 2002. WSEAS, pp 147–153

Castelfranchi C (1998) Modelling social action for AI agents. Artif Intell 103(1–2):157–182

Castelfranchi C, Conte R, Paolucci M (1998) Normative reputation and the costs of compliance. J Artif Soc Soc Simul 1(3). http://jasss.soc.surrey.ac.uk/1/3/3.html

Conte R (1999) Social intelligence among autonomous agents. Comput Math Organ Theory 5: 202–228

Conte R, Castelfranchi C (1995) Cognitive and social action. UCL Press, London

Conte R, Paolucci M (1999) Reproduction of normative agents: a simulation study. Adapt Behav 7 (3/4):301–322

Conte R, Paolucci M (2002) Reputation in artificial societies: social beliefs for social order. Kluwer, Dordrecht

Conte R, Paolucci M (2003) Social cognitive factors of unfair ratings in reputation reporting systems. In: Liu J, Liu C, Klush M, Zhong N, Cercone N (eds) Proceedings of the IEEE/WIC international conference on web intelligence – WI 2003. IEEE Computer Press, Halifax, pp 316–322

Conte R, Castelfranchi C, Dignum F (1998) Autonomous norm-acceptance. In: Müller JP, Singh MP, Rao AS (eds) ATAL '98: proceedings of the 5th international workshop on intelligent agents V, agent theories, architectures, and languages. Springer, London, pp 99–112

Cravens K, Goad Oliver E, Ramamoorti S (2003) The reputation index: measuring and managing corporate reputation. Eur Manage J 21(2):201–212

Dellarocas CN (2003) The digitalization of word-of-mouth: promise and challenges of online feedback mechanisms (MIT Sloan working paper, 4296–03). MIT Sloan School of Management, Cambridge, MA

Di Tosto G, Giardini F, Conte R (2010) Reputation and economic performance in industrial districts: modelling social complexity through multi-agent systems. In: Takadama K, Cioffi-Revilla C, Deffuant G (eds) Simulating interacting agents and social phenomena: the second world congress, vol 7, ABSS. Springer, Tokyo, pp 165–176

Dunbar R (1996) Grooming, gossip, and the evolution of language. Faber and Faber, London

Fehr E, Gachter S (2000) Fairness and retaliation: the economics of reciprocity. J Econ Perspect 14:159–181

Fioretti G (2005) Agent based models of industrial clusters and districts (Working paper, series urban/regional, 0504009). EconWPA http://ideas.repec.org/p/wpa/wuwpur/0504009.html

Frith CD, Frith U (2006) How we predict what other people are going to do. Brain Res 1079 (1):36–46

Giardini F, Cecconi F (2010) Social evaluations, innovation and networks. In: Ahrweiler P (ed) Innovation in complex social systems. Routledge, London, pp 277–289

Giardini F, Di Tosto G, Conte R (2008) A model for simulating reputation dynamics in industrial districts. Simul Model Pract Theory 16:231–241

Gintis H, Smith EA, Bowles S (2001) Costly signaling and cooperation. J Theor Biol 213:103–119

Gluckman M (1963) Gossip and scandal. Curr Anthropol 4:307–316

Gray ER, Balmer JMT (1998) Managing corporate image and corporate reputation. Long Rage Plann 31(5):695–702
Greif A (1993) Contract enforceability and economic institutions in early trade: the maghribi traders' coalition. Am Econ Rev 83(3):525–548
Hales D (2002) Group reputation supports beneficent norms. J Artif Soc Soc Simul 5(4). http://jasss.soc.surrey.ac.uk/5/4/4.html
Huynh D, Jennings NR, Shadbolt NR (2004) Developing an integrated trust and reputation model for open multi-agent systems. In: Falcone R, Barber S, Sabater-Mir J, Singh M (eds) Proceedings of the 7th international workshop on trust in agent societies, New York, 19 July 2004, pp 65–74
IC3 (2005): IC3 2005 Internet crime report http://www.ic3.gov/media/annualreport/2005_IC3Report.pdf
Jøsang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. Decis Support Syst 43(2):618–644
Koenig S, Hudert S, Eymann T, Paolucci M (2008) Towards reputation enhanced electronic negotiations for service oriented computing. In: Falcone R, Barber S, Sabater-Mir J, Singh M (eds) Trust in agent societies: 11th international workshop, TRUST 2008, Estoril, 12–13 May 2008, revised selected and invited papers. Springer, Berlin, pp 273–291
Kreps D, Wilson R (1982) Reputation and imperfect information. J Econ Theory 27:253–279
Leslie AM (1992) Autism and the 'theory of mind' module. Curr Dir Psychol Sci 1:18–21
Marsh S (1992) Trust in distributed artificial intelligence. In: Castelfranchi C, Werner E (eds) Artificial social systems, 4th European workshop on modelling autonomous agents in a multi-agent world, MAAMAW '92, S. Martino al Cimino, 29–31 July 1992, selected papers. Springer, Berlin, pp 94–112
Marsh S (1994a) Formalising trust as a computational concept. PhD Thesis, Department of Computing Science and Mathematics, University of Stirling, Stirling. http://www.cs.stir.ac.uk/research/publications/techreps/pdf/TR133.pdf
Marsh S (1994b) Optimism and pessimism in trust. In: Proceedings of the Ibero-American conference on artificial intelligence (IBERAMIA-94), Caracas, 25–28 Oct 1994
Melnik MI, Alm J (2002) Does a seller's eCommerce reputation matter? Evidence from eBay auctions. J Ind Econ 50(3):337–349
Miceli M, Castelfranchi C (2000) The role of evaluation in cognition and social interaction. In: Dautenhahn K (ed) Human cognition and social agent technology. Benjamins, Amsterdam, pp 225–262, chapter 9
Moukas A, Zacharia G, Maes P (1999) Amalthaea and histos: multiAgent systems for WWW sites and reputation recommendations. In: Klusch M (ed) Intelligent information agents: agent-based information discovery and management on the internet. Springer, Berlin, pp 292–322
Noon M, Delbridge R (1993) News from behind my hand: gossip in organizations. Organ Stud 14: 23–36
Nowak MA, Sigmund K (1998a) Evolution of indirect reciprocity by image scoring. Nature 393: 573–577
Nowak MA, Sigmund K (1998b) The dynamics of indirect reciprocity. J Theor Biol 194:561–574
Paine R (1967) What is gossip about? An alternative hypothesis. Man 2(2):278–285
Panchanathan K, Boyd R (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature 432:499–502
Paolucci M (2000) False reputation in social control. Adv Complex Syst 3(1–4):39–51
Paolucci M (ed) (2009) Proceedings of the first international conference on reputation: theory and technology – ICORE 09, Gargonza. Institute of Science and Technology of Cognition, National Research Centre (ISTC-CNR), Rome. http://pagesperso-systeme.lip6.fr/Erika.Rosas/pdf/ICORE09.pdf
Paolucci M, Conte R (2009) Reputation: social transmission for partner selection. In: Trajkovski GP, Collins SG (eds) Handbook of research on agent-based societies: social and cultural interactions. IGI, Hershey, pp 243–260

Paolucci M et al (2009) Theory and technology of reputation (Technical report, FP6 Research Project "Social knowledge for e-Governance" (eRep)). Institute of Science and Technology of Cognition, National Research Centre (ISTC-CNR), Rome. http://issuu.com/mario.paolucci/docs/erep_booklet

Pinyol I, Paolucci M, Sabater-Mir J, Conte R (2008) Beyond accuracy: reputation for partner selection with lies and retaliation. In: Antunes L, Paolucci M, Norling E (eds) Multi-agent-based simulation VIII: international workshop, MABS 2007, Honolulu, 15 May 2007, revised and invited papers (Lecture notes in computer science 5003). Springer, Berlin, pp 128–140

Porter M (1998) Clusters and the new economics of competition. Harv Bus Rev 76:77–90

Ramchurn SD, Huynh D, Jennings NR (2004a) Trust in multiagent systems. Knowl Eng Rev 19(1):1–25

Ramchurn SD, Sierra C, Godo L, Jennings NR (2004b) Devising a trust model for multi-agent interactions using confidence and reputation. Int J Appl Artif Intell 18:833–852

Rasmusson L (1996) Socially controlled global agent systems (Working paper). Department of Computer and Systems Science, Royal Institute of Technology, Stockholm

Rasmusson L, Janson S (1996) Simulated social control for secure internet commerce. In: Proceedings of the 1996 workshop on new security Paradigms, Lake Arrowhead. ACM Press, New York, pp 18–25

Regan K, Cohen R (2005) Indirect reputation assessment for adaptive buying agents in electronic markets. In: Proceedings of workshop business agents and the semantic web (BASeWEB'05), 8 May 2005, Victoria, British Columbia

Resnick P, Zeckhauser R (2002) Trust among strangers in internet transactions: empirical analysis of eBay's reputation system. In: Baye MR (ed) The economics of the internet and e-commerce, vol 11, Advances in applied microeconomics. Elsevier Science, Amsterdam, pp 127–157

Ripperger T (1998) Ökonomik des Vertrauens: Analyse eines Organisationsprinzips. Mohr Siebeck, Tübingen

Rocco E, Warglien M (1995) La comunicazione mediata da computer e l'emergere dell'opportunismo elettronico. Sistemi Intelligenti 7(3):393–420

Rose C, Thomsen S (2004) The impact of corporate reputation on performance: some Danish evidence. Eur Manage J 22(2):201–210

Saam N, Harrer A (1999) Simulating norms, social inequality, and functional change in artificial societies. J Artif Soc Soc Simul 2(1), http://jasss.soc.surrey.ac.uk/2/1/2.html

Sabater J, Paolucci M (2007) On representation and aggregation of social evaluations in computational trust and reputation models. Int J Approx Reason 46(3):458–483

Sabater J, Sierra C (2002) Reputation and social network analysis in multi-agent systems. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS 2002), 15–19 July 2002, Bologna. ACM Press, New York, pp 475–482

Sabater J, Sierra C (2005) Review on computational trust and reputation models. Artif Intell Rev 24(1):33–60

Sabater J, Paolucci M, Conte R (2006) REPAGE: REPutation and ImAGE among limited autonomous partners. J Artif Soc Soc Simul 9(2), http://jasss.soc.surrey.ac.uk/9/2/3.html

Schillo M (1999) Vertrauen: Ein Mechanismus zur sicheren Koalitionsbildung in künstlichen Gesellschaften (Trust: a mechanism for reliable coalition formation in artificial societies). Master's thesis, Department of Computer Science, Saarland University, Saarbrücken, Germany.

Schillo M, Funk P, Rovatsos M (2000) Using trust for detecting deceitful agents in artificial societies. Appl Artif Intell 14:825–848

Sell J, Wilson R (1991) Levels of information and contributions to public goods. Soc Forces 70: 107–124

Sommerfeld RD, Krambeck H, Milinski M (2008) Multiple gossip statements and their effects on reputation and trustworthiness. Proc R Soc B 275:2529–2536

Squazzoni F, Boero R (2002) Economic performance, inter-firm relations and local institutional engineering in a computational prototype of industrial districts. J Artif Soc Soc Simul 5(1). http://jasss.soc.surrey.ac.uk/5/1/1.html

Utz S, Matzat U, Snijders CCP (2009) On-line reputation systems : the effects of feedback comments and reactions on building and rebuilding trust in on-line auctions. International Journal of Electronic Commerce 13(3):95–118

Wilson DS, Wilczynski C, Wells A, Weiser L (2000) Gossip and other aspects of language as group-level adaptations. In: Heyes C, Huber L (eds) The evolution of cognition. MIT Press, Cambridge, pp 347–366

Yamagishi T, Matsuda M, Yoshikai N, Takahashi H, Usui Y (2009) Solving the lemons problem with reputation. An experimental study of online trading. In: Cook KS, Snijders C, Vincent B, Cheshire C (eds) eTrust: forming relationships in the online world. Russel Sage Foundation, New York, pp 73–108

Yu B, Singh MP (2000) A social mechanism of reputation management in electronic communities. In: Klusch M, Kerschberg L (eds) Cooperative information agents IV, the future of information agents in cyberspace, 4th international workshop, CIA 2000, Boston, 7–9 July 2000, proceedings (Lecture notes in computer science, 1860). Springer, Berlin, pp 154–165

Zacharia G (1999) Trust management through reputation mechanisms. In: Castelfranchi C, Falcone R, Firozabadi BS (eds) Proceedings of the workshop on deception, fraud and trust in agent societies at autonomous agents'99, Seattle. pp 163–167

Zacharia G, Moukas A, Maes P (1999) Collaborative reputation mechanisms in electronic marketplaces. In: Proceedings of the 32nd annual Hawaii international conference on system sciences (HICSS-32), 5–8 Jan 1999, Maui, Track 8: software technology. IEEE computer society, Los Alamitos