Reputation for complex societies

Giardini, Francesca; Conte, Rosaria; Paolucci, Mario

Published in:
Simulating Social Complexity

DOI:
10.1007/978-3-319-66948-9_18

Chapter 18
Reputation for Complex Societies

Francesca Giardini, Rosaria Conte, and Mario Paolucci

Abstract Reputation, the germ of gossip, is addressed in this chapter as a distributed instrument for social order. In literature, reputation is shown to promote (a) social control in cooperative contexts—like social groups and subgroups—and (b) partner selection in competitive ones, like (e-) markets and industrial districts. Current technology that affects, employs and extends reputation, applied to electronic markets or multi-agent systems, is discussed in light of its theoretical background. In order to compare reputation systems with their original analogue, a social cognitive model of reputation is presented. The application of the model to the theoretical study of norm-abiding behaviour and partner selection are discussed, as well as the refinement and improvement of current reputation technology. The chapter concludes with remarks and ideas for future research.

18.1 Reputation in Social Systems: A General Introduction

Ever since hominid settlements started to grow, human societies have needed to cope with problems of social order. 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"?

Social knowledge like reputation and its transmission (i.e. gossip) play a fundamental role in creating and maintaining 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 even unwilling and unaware individuals derive from the generation, transmission and manipulation of a special type of social belief, which

F. Giardini (✉)
Faculty of Behavioral and Social Sciences, Department of Sociology,
University of Groningen, Groningen, The Netherlands (NL)
e-mail: f.giardini@rug.nl

R. Conte • M. Paolucci
Laboratory for Agent-Based Social Simulation (LABSS), Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

B. Edmonds, R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-66948-9_18

has contributed to the regulation of natural societies from the dawn of mankind (Dunbar 1996). People use reputational information for many things, including: to make decisions about possible interactions, to evaluate candidate partners and to understand and predict their behaviours (Alexander 1987).

It has long been known that reputation is a fundamental generator, vehicle and manipulator of social knowledge for enforcing reciprocity and other social norms (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 since the early 1980s (Kreps and Wilson 1982). However, at that stage, there was little understanding of its dynamic and cognitive underpinnings. Despite its critical role in the enforcement of altruism, cooperation and social exchange, the socio-cognitive study of reputation is relatively new. Hence, how this critical type of knowledge is manipulated in the minds of agents, how social structures and infrastructures generate, transmit and transform it, has not yet been fully clarified. Consequently, the full picture of how it affects agents' behaviour is also unclear. Partly, this is because reputation extends beyond the boundaries of academic disciplines, emerging as a prototypical cross-disciplinary topic (Paolucci and Sichman 2014).

The aim of this chapter is to guide the reader through the multiplicity of computational approaches concerned with the reputation mechanism 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. We will follow this with three different applications of the cognitive theory of reputation to model social phenomena: the Sim-Norm model, the Socrate framework and the Repage architecture.

This introduction will be followed. In order to lay the ground for understanding the multiplicity of reputation, we will present by an outline of reputation research in some different domains, namely, social psychology, management and experimental economics and agent-based simulation. This will show the variety of viewpoints that can be used to describe and explore this complex phenomenon. We will then focus on some of the work in electronic markets and multi-agent simulations that include reputation mechanisms. 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 has contributed to a growing number of auction sites that facilitate the exchange of goods between individual consumers, without guaranteeing either transparency or the safety of the 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. 18.6 we propose a cognitive model of reputation, which aims to solve some of the problems left open by existing systems, starting 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 is of the model and their results are then provided. We also describe a set of simulation studies on gossip, in which private transmission of unverified

information is able to support cooperation in a public goods game. Moving from the observation that reputation and punishment are considered the most important mechanisms for social control, a systematic comparison of their effects will show that their combination represents a powerful way of detecting and deterring cheaters. Finally, we draw some conclusions about some future directions for research in this area.

18.2 State of the Art: An Overview on Reputation in Natural and Artificial Societies

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 potential partner(s) in an interaction and thus to form expectations about 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 develop. Reputation is therefore a coordination device whose predictive power is essential in social interactions (Paolucci and Conte 2009).

Reputation and its transmission (gossip) have an extraordinary preventive power: it substitutes personal experience in (a) identifying cheaters and isolating them and in (b) finding trustful partners. It makes available most of the benefits of evaluating someone, without 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 members within the group (Barkow 1996) while, at the same time, isolating those in out-groups. Gossip contributes to stratification and social control, since it works as a tool for sanctioning deviant behaviours and for promoting those behaviours that are functional with respect to the group's goals and objectives (e.g. via a learning process). Reputation is also considered as a means for sustaining and promoting the diffusion of norms and norm conformity (Wilson et al. 2000). On the other hand, reputation can be used to pursuit self-interest, either by promoting one's achievements or by spreading negative information about others (Paine 1967; Noon and Delbridge 1993).

Reputation plays a key role in evolutionary theories of cooperation, supporting indirect reciprocity (Nowak and Sigmund 1998a, b, 2005). Theories of indirect reciprocity explain large-scale human cooperation in terms of conditional helping by individuals who want to uphold a reputation and then to be included in future cooperation (Panchanathan and Boyd 2004). By means of computer simulations, Nowak and Sigmund (1998a, b) showed that reputation can sustain the emergence of indirect reciprocity—getting people to cooperate (even with strangers) in order to receive cooperation, without the necessity of any kind of contract or keeping

track of contributions. Theories of indirect reciprocity explain large-scale human cooperation in terms of conditional helping by individuals who want to uphold a reputation and then to be included in future cooperation (Panchanathan and Boyd 2004). In a "market for cooperators" (Noë and Hammerstein 1994), or in partner choice, building a positive reputation for generosity can be seen as a long-term investment. Here, individuals may compete for the most altruistic partners, leading non-altruists to become ostracized (Roberts 1998).

As Alexander (1987) pointed out "indirect reciprocity involves reputation and status, and results in everyone in the group continually being assessed and reassessed". In the last few years, attention to reputation has grown both within single disciplines and in interdisciplinary contexts (Milinski 2016; Wu et al. 2016). This has involved a variety of methodologies, going from online large-scale experimental studies using dynamic networks (Rand et al. 2011; Wang et al. 2012) to economic laboratory experiments, and has included important advances in the study of reputation as a means to support cooperation in a variety of contexts (Beersma and Van Kleef 2011; Piazza and Bering 2008; Sommerfeld et al. 2008).

Reputation and gossip are also crucial in other fields of the social sciences like management and organization science, governance and business ethics, where the importance of reputation in branding became apparent (Fombrun and Shanley 1990). The economic interest in the subject matter came from the fact that reputation can be applied at 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 (Fombrun 1996). Rose and Thomsen (2004) claim that a good reputation and a good financial performance are mutually dependent—a good reputation may influence the financial assets 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 CEOs 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, and is thus similar to individual perception, whereas the reputation results more from the firm's communication and long-term strategy. Generally speaking, corporate reputation is treated as an aggregate evaluation that stakeholders, consumers, managers, employees and institutions form about a firm. However, the mechanisms leading to the final result are not well defined.

If social order is a constant of human evolution, it is particularly crucial in an e-society where the boundaries of interaction are widening. The increasingly fast development of ICT technologies dramatically enlarges the range of interaction among users, generating new types of aggregation, from civic communities to electronic markets and from professional networking to e-citizenship. 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 (e.g. 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). Detailed studies of the effects of technological infrastructures on interaction styles and modes are lacking. Perhaps, an exception to this is represented by the research on the effects of asymmetry of information within the markets. Asymmetry of information is known to encourage fraud and low-quality production in many situations. As exemplified by Akerlof (1970), asymmetry of information can drive 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 such a 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. This phenomenon is an intrinsic feature of e-markets, but goes back to eleventh-century Maghribi traders moving along the coast of the Mediterranean Sea (Greif 1993). Contemporary online traders such as users of Internet auction sites face the same problem of mediaeval traders: online buyers can learn about the quality (or condition) of the good only once they have already paid for it.

Auction sites vary from the very generic, concerning the products being offered and operated on a global scale (e.g. eBay), to those that focus on specific products on a national scale (many car auction sites). Buying through auction sites offers less control to the buyers than even online retailers, 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. Sellers on the Internet may actively try to communicate their reputation to potential buyers, increasing the expected impact of reputation on buying decisions. 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. (2004) show that reputation has a significant positive effect on the quality of products. In any case, the strength of reputational mechanisms does not seem to be diminished by the spread of anonymous contexts in which interactions take place at a distance and are mediated by a computer (as happens online). In this sense, the new technologies allow for more information. A greater number of people can interact due to overcoming spatial limitations. These new opportunities for large-scale interaction, as well as the chance to make opinions accessible to the community of Internet users (i.e. bidirectionality), have allowed the development of systems based on online feedback mechanisms (Dellarocas 2003). We are witnessing the proliferation of services that rely on reputation systems (e.g. eBay, Amazon, TripAdvisor), and experimental studies show that even in online

anonymous contexts, where the way in which reputation is assigned is opaque, reputation is used to actively avoid defectors (Capraro et al. 2016). Technical challenges for large-scale systems can be met with the use of simple reputation systems, as in the case of collaborative filtering algorithms (Petroni et al. 2016).

Despite the role of reputation in economic transactions, online reputation systems are only moderately efficient (Bolton et al. 2004; Resnick and Zeckhauser 2001), showing a stronger effect of negative feedbacks on price reduction than a positive one on price increase (Diekmann et al. 2014).

In all of these cases, 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 resultant reputation score. This mechanism is debatable alone and can be insufficient, but it can be complemented by detailed comments or forums. For example, when forums are available, this is the solution chosen by TripAdvisor, whose users can provide detailed comments about hotels, restaurants, tourist attractions and services. These comments are displayed along with real reputation exchanges that are performed in parallel, thus offering interested users as much information as possible. Moreover, many people do not bother to provide reputational feedback (under-provision), and if they do, they lean on providing only positive reports (overscoring).

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-cost; (2) reputation has more impact than directly acquired information. In a simulation study, Pinyol et al. (2008) showed that if agents transmitted only their own evaluations about one another (image), the circulation of social knowledge ceases quickly. 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. In a simulation study, Pinyol et al. (2008) showed that if agents transmitted only their own evaluations about one another (image), the circulation of social knowledge ceases quickly. 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.

18.3 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 prosocial behaviour—be it cooperative, altruistic or norm abiding—among autonomous, i.e. self-interested agents. Thanks to computational methods, social simulation has contributed to our understanding

of reputation as a means to promote norm-abiding behaviour in social groups and as a tool for improving partner selection in electronic markets and computational settings.

Several attempts have been made to model and use reputation in artificial societies, especially in two subfields of information technology: 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 (Moukas et al. 1999; Zacharia 1999; Zacharia et al. 1999). 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.

18.3.1 Online Reputation Reporting Systems

The continuously growing volume of transactions on the World Wide Web and the growing number of frauds that appears to entail[1] have led scholars from different disciplines to develop new online reputation reporting systems. These systems are intended 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 subsets, 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 one 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 against other 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, an action that may be punished by a "reputation decline". These individual-based models (Bachmann 1998; Marsh 1994a; Ripperger 1998) differ with regard to their memory span. Agents may forget their experiences slowly, fast or never, and this has important consequences for the dynamics of the overall level of trust in the system.

[1]According to the US-based Internet Crime Complaint Center (IC3), losses as a result of auto-auction fraud exceeded $8.2 million dollars in 2011.

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 1997; Rasmusson 1996; Rasmusson and Janson 1996;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 use upon different methods to weigh the third-party information and to deal with "friends of friends". Thus the question arises as to how to react to information from agents who do not seem to be very trustworthy. A similar problem arises with 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.

Regan and Cohen (2005) propose a system for computing indirect and direct reputation within 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 reputation in the face of undesirable practices 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 the model by Regan and Cohen and similar other models concerns the transmission of reputation. In these kinds of models, agents only react to reputation requests, while proactive, spontaneous delivery of reputation information to selected recipients is not considered. However, this simple solution is quite effective. On the other hand, despite its simplicity, these types of model tackle the problem of collusion between rating agents, because by keeping the evaluation of sellers remains among buyers (i.e. not disclosing it to the sellers). Therefore sellers cannot influence their own scores.

Turning to electronic marketplaces, classic systems like eBay show a characteristic bias towards positive evaluations (Resnick and Zeckhauser 2002). This suggests that factual cooperation among users at the information level may lead to a "courtesy" equilibrium (Conte and Paolucci 2003). As Cabral and Hortaçsu (2010) formally prove, 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 but enjoy a good reputation: after a good start, they will hardly receive negative feedback and even if they do, it will not get to the point of spoiling their good name. Under such conditions, even good sellers may have an incentive to sell lemons, considering that it takes time for their reputation scores to go down.

Intuitively, the courtesy equilibrium reduces the deterrent effect of reputation. If a reputation system is meant to reduce 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.

18.3.2 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. 2004; Sabater and Sierra 2004; Pinyol and Sabater-Mir 2013) present interesting ideas and advances over conventional online reputation systems, with their notion of a distributed 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 link-by-link through the network of agents. It also has a mechanism to allow agents to include other agents' testimonies in its reputation calculations. Agents store information about the outcome of every transaction they ever 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 a partial view instead of global score 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 cooperating agents experienced a slow, almost linear increase in reputation. However, some problems remain. The model does not allow agents to combine their own experience with the network information. Thus, it might take unnecessarily long to react to a suddenly defecting agent that cooperated before. In addition, Yu and Singh do not give an explanation of how their agent-centred storage of social knowledge (e.g. the referral chains) is supposed to be organized. Consequently, no analysis of network load and storage intensity can be done.

ReGreT (Sabater 2004) is another MAS application in which the link between trust and reputation is very strong. In this, reputation is only one of the dimensions an agent resort to in order to evaluate the trustworthiness of another agent. In ReGreT, reputational information and direct experience have different values, and the former is considered less reliable than the latter.

A system called Liar Identification for Agent Reputation (LIAR) has been proposed by Muller and Vercouter (2008), based on three levels of reputation: direct, indirect and recommendation based. To implement those elements, LIAR explicitly models a social commitment mechanism, social norms and the operations over them.

SOARI (Service Oriented Architecture for Reputation Interaction) is a reputation ontology that has been proposed by Nardin et al. (2008). SOARI is a service-oriented architecture that provides support to the semantic interoperability among agents that implement heterogeneous reputation models. The main contribute of SOARI is to provide a mapping among different reputation models, represented by a common reputation ontology especially designed for agents' interaction, in the form of a service that can be executed externally to agents and is available online as an on-demand service for agents.

As these example shows, the "agentized environment" produces interesting solutions that may apply also to online communities. This is for two main reasons. *Firstly*, in this environment two problems of order arise: meeting users' expectations (external efficiency) and promoting 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. In order to promote the former, agents must be in an environment where they evaluate and act upon each other's behaviours. *Secondly*, agent systems can be used to help determine (a) what type of agents, (b) what type of beliefs and (c) what type of processes among agents are required to achieve useful social control. More specifically, they can be used to map out what type of agent and processes are needed for *which* desirable result, including better efficiency, encouraging equity (and hence users' trust), discouraging discrimination and fostering collaboration at the information level or object level (or both).

However, in models of Internet systems, 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 their global reputation value. The solutions proposed for MAS systems are interesting, but these are insufficient to meet the problems left open by online contexts. There is a tendency to consider reputation as an external attribute of agents without taking into account the processes of creation and transmission of that reputation. Is there an alternative? How can we understand the effects of reputation on transactions if we do not model the process of reputation creation and transmission?

18.4 An Alternative Approach: The Social Cognitive Process of Reputation

Current models operate with a highly simplified model of reputation, in which different experiences and items of information are reduced to a single accumulator. In this section, we will model reputation as a social cognitive process and briefly discuss advantages and disadvantages of this approach.

A social cognitive process involves symbolic mental representations (e.g. social beliefs and goals[2]) that are manipulated by individuals and agents in the process of social reasoning by means of the operations that agents perform upon them (social reasoning).

Social cognitive processes are aimed at modelling (and possibly implementing) systems acting in a social—be it natural or artificial—environment. These processes employ explicit representations of a 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 thus triggering the processes of (social learning, influence, and control). To represent reputation as a social cognitive process, two different constructs are needed, namely, image and reputation. After giving the definition of those constructs, we will show how agents can behave when evaluating someone and transmitting these evaluations. Thus playing one of three different roles: evaluator, beneficiary and target.

18.4.1 Image and Reputation

An *image* consists of a set of evaluative beliefs (Miceli and Castelfranchi 2000) held by an agent (the "evaluator") concerning the characteristics of another agent (the "target"). It is an assessment of its positive or negative qualities with regard to a norm or competence. 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.

An agent's *reputation* we argue is distinct from, although closely related to, its image. More precisely, we define *reputation* as consisting of three distinct but interrelated objects: (1) a cognitive representation, i.e. a believed evaluation of another agent; (2) a population object, i.e. an evaluation that is propagated to others; and (3) an objective emergent property at the agent level, i.e. what the agent is believed to be. As an illustration, when we say that "John has a very good reputation as a dentist", we are implicitly assuming that (1) someone believes that he is good at his job, (2) an indefinite number of people share that belief, and (3) he actually possesses some skills; therefore his reputation is grounded in some objective properties.

Reputation is a highly dynamic phenomenon in two distinct senses: it is subject to change, especially due to the effect of corruption, errors and deception, and it emerges as an effect of a multilevel process within the society of agents (Conte and Paolucci 2002). This involves emergence both from agents to society and from

[2]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).

society back to the individual agents. 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. Once it reaches the population level, it gives rise to an additional property at the agent level. Reputation is the immaterial, more powerful equivalent of the scarlet letter sewn to one's clothes Nathaniel Hawthorne described in his masterpiece. It is more powerful because it may not be perceived by the individual to whom it is attached and therefore harder for an individual (him/her) to control or manipulate. The objective nature of reputation (in our sense) also makes it impersonal, and therefore, spreading reputation can carry less responsibility than spreading image.

To formalize these concepts, we will begin by defining the building blocks of "image". An agent has made an evaluation when he or she believes that a given entity, be it another agent, an organization, a firm, etc., can achieve a specific goal of some agent who is (often, but not always) the same as the evaluator. An agent has made a *social evaluation when his or her belief concerns another agent as a means for achieving this goal. Thus*, E targets T and benefits B. 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.

To make this analysis more concrete, we will start with an example in which we consider a classic multi-agent situation in which a set of agents fight 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). For reasons of simplicity, the agents carrying out the evaluation are restricted to the norm followers: $E = N = N \cup C$: indeed, if normative agents benefit globally from the presence of the norm, cheaters in this simple setting benefit even more; 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 (E, T and B) coincide. General behavioural norms, such as "Do not commit murder", apply to, benefit, and are evaluated by all agents. However, there are also 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 programmes for children, broadcast in the afternoon. Here, we can identify three more or less distinct sets. The 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 programmes and the decision-makers at the broadcast stations. There may be a non-empty intersection

between E and T but no full overlap. If the target of evaluation is the broadcaster itself (a supra-individual entity), the intersection can be considered to be empty.

Extending this formalization to include reputation, we have to differentiate it 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. While image is based on direct experience or observation, therefore an evaluator is assumed to believe his/her own evaluation; reputation therefore involves one more set of agents: in addition to evaluators E, targets T and beneficiaries B, we have a set M of memetic agents who share the meta-belief. This means that they simply believe that some other agents had a positive experience with John, therefore they hold the meta-belief that John has a positive reputation ("I believe that others believe that he is a good dentist"). It is important to stress the fact that a memetic agent does not need to hold the evaluation belief, but she simply need to transmit it. If they contribute to the diffusion of reputation, the memetic agents can also be labelled as gossipers G. 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 non-empty.

18.4.2 Identifying Reputational Roles

We have seen that agents may play more than one role simultaneously: evaluator, beneficiary, target and memetic/gossiper. In order to implement a socio-cognitive model of reputation, we need to describe the characteristics of the four roles in more detail.

18.4.2.1 Evaluator

Autonomous agents continually asses their environment and form evaluations as effect of interaction and perception. Social evaluations are formed when agents evaluate one another with regard to their goals (Castelfranchi 1998).

This image, based on direct experience, drives future actions: it serves to identify friends and to avoid enemies or cheating partners. Agents also observe interactions between third parties and evaluate them with regard to the goals or interests of a given set of agents (the beneficiaries). Information thus obtained may be used to draw inferences about the target's likelihood to violate other rights in the future. 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, organizations or abstract social entities.

18.4.2.2 Beneficiary

A beneficiary is the entity that benefits from the action with regard to which targets are evaluated. Beneficiaries can be individual agents, groups and organizations 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.

18.4.2.3 Target

The target of social evaluation is the entity that is evaluated. Targets of reputation (targets) should be autonomous agents endowed with mental states, possibly with an explicit decision-making or deliberative capacity. Consequently, they are a locus of social responsibility: they hold the power to prevent social harm and possibly to respond for it, in case any harm occurs.

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.

18.4.2.4 Gossiper (Memetic Agent)

An agent is a (potential) memetic agent 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, memetic agents do not necessarily share the corresponding image (social evaluation) of the target. That is, they do not necessarily believe it to be true.

Memetic agents (if they are also targets) may deserve a negative evaluation; they may actually convey information that they hold to be false in order to enjoy the advantages of sharing reputation information. By sharing reputation, the agent will be considered as part of the in-group by other evaluators, and therefore gain a good reputation without sustaining the costs of its acquisition.

18.5 Implementing the Social and Cognitive Processes of Reputation: Sim-Norm and Repage

Sim-Norm was the first attempt to implement the social and cognitive theory of reputation. The model was developed to examine the effect of reputation on the efficiency of a norm of precedence (Conte et al. 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, and its results confirmed that reputation can have a positive impact on social control.

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:

1. Norm-based social order can be maintained, and its costs reduced via distributed social control.
2. 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. 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).

The model defines agents as objects moving in a two-dimensional environment (a 10×10 grid) with randomly scattered food. At the beginning of each 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 of the simulation, 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", that is, 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 characterized 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). The first two series of experiments showed that normative agents perform less well than nonnormative 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. 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 cooperative agents with the capacity to exchange with their (believed-to-be) respectful neighbours at distance one from them 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. 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 Sim-Norm to investigate the effects of group reputation. In his model, agents

are given the cognitive capacity to categorize other agents as members of a group and project reputation onto whole groups instead of individual agents (a form of stereotyping).

Repage (Sabater et al. 2006) is a computational system for reputation management. Based on a model of reputation, image and their interplay, Repage provides evaluations of potential partners and is fed with information transmitted 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.

In order to preserve their autonomy, agents need to *decide* whether 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; 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 specialized subsystem. This is what Repage provides.

Repage is a sophisticated cognitive architecture that operates on a subset of the predicates that constitute the memory of the agent, that is, of those predicates that are relevant for dealing with image and reputation. Predicates about reputation, as discussed above, must contain an evaluation about a target which contains three aspects: the type of the evaluation (either personal experience or image or third party image), the role of the target (either informant or seller) and the actual content. To store the content, a simple number is used, as in eBay 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 tells that "People are saying that Jane is 0.234 good". To capture the lack of precision coming from vague utterances, e.g. "I believe that agent X is good, I mean, very good — good, that is", and from noise in the communication or in the recollection from memory, the actual value of an evaluation is represented in a fuzzy way, by a n-tuple of positive real values that sum to one.

Finally, each predicate has a strength value associated to it. This value is a function of the strength of its antecedents and of some special characteristics intrinsic to that type of predicate. The network of dependencies specifies which predicates contribute to the values of other predicates. In fact, each predicate in the Repage memory has a set of antecedents and a set of consequences. If an antecedent

changes its value or is removed, the predicate is notified, thanks to the work of the detectors. Then the predicate recalculates its value and notifies the change to its consequences. Aggregation and other interesting properties of these representations are detailed in Sabater and Paolucci (2007). An example of Repage in action can be found in Quattrociocchi et al. (2008).

To illustrate the behaviour of Repage, let us consider an example about a potential purchase. The 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 forums, auctions and in most agent systems. Then, 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 an informer, he resorts to a default image (i.e. usually quite low). The uncertain image as an informer adds uncertainty to the value of the communication, resulting in a decision to look for more information.

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 the fuzzy metric presented above) 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 coherent, 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. Consequently, 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. As a final wave of feedback, the image of the three agents who gave a good evaluation of *Z* as an informer is increased, while the image of the other three is decreased. This feedback is a necessary and relevant part of the Repage model.

Taking into account the correlations between different reputation attributes, Nardin et al. (2014) compare Repage with other architectures via a multivariate statistical approach. Their analysis shows that, in most cases, there is a benefit in using a more expressive communication language.

18.5.1 A Simulation Model of Reputation Spreading in an Industrial District: SOCRATE

SOCRATE is an attempt to test the cognitive theory of reputation in an ideal-typical economic setting, modelled 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 the interplay between the market structure and the social relationships among agents. Social links and the resulting social structure, usually informal, are defining features of industrial clusters (Porter 1998; Fioretti 2005; Squazzoni and Boero 2002), in which trust and reputation play a crucial role. 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 and to exert social control on cheaters. We designed 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 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.

This model was developed with the aim of answering the following questions: 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?

There are two different kinds of interactions among agents in the model: 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. The latter consists in the flow of social evaluations among the firms, which is of paramount importance in this setting, where 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. We implemented three layers: Layer 0 (*L0*) is represented by leader firms that supply the final product and are supplied by firms on Layer 1 (*L1*). On Layer 2 (*L2*), there are firms providing raw material to firms in *L1*.

Reputation and image transmission are exchanged within layers, so for instance firms on *L0* and *L1* are not allowed to talk each other. 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 cannot 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 a table in which all the values of the tested partners are recorded and stored for future selections. Under 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.

Our results showed that the quality of products was higher in the cluster with reputational information, compared to the cluster with image, for the same percentages of cheating. We also replicated the results by varying the distribution of firms on the three layers, thus designing a market with harsh competition for good partners, and we found that the exchange of reputational information also allows the whole cluster to obtain higher profits (Di Tosto et al. 2010).

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.

18.6 Gossip as Reputation Transmission and Its Effect on Cooperation in Social Dilemmas

Gossip is a multifaceted social phenomenon, widespread in human societies and serving several functions: it is a valuable source of information about community and its members, but it is also essential to map the social environment, to promote membership and to sanction deviant behaviours in a public way (Giardini and Conte 2012). In human groups, exchanging evaluations serves as a means to create and maintain relationships between individuals, and it might be pivotal to either the creation or the enforcement of other kinds of relationships (friendship, acquaintances, business, etc.).

When cooperation is framed as a public goods game (Hardin 1968; Gardner, Ostrom, Walker 1990), cooperation can emerge only if individuals sacrifice short-term gains in favour of the long-term collective good. In large groups, this translates into a high probability that individuals will tend to interact with complete strangers with little or no opportunities for positive reciprocity. Simulation data and lab experiments show that cooperation can hardly be sustained in groups, unless costly punishment is provided (Carpenter 2007; Fehr and Gachter 2000). Although effective in many contexts, costly punishment increases the amount of cooperation but not the average pay-off of the group (Dreber et al. 2008). Those who punish pay a cost for that. In repeated games, cooperators who do not bear the costs of punishing defectors are better off than cooperators who punish (Ohtsuki et al. 2009). Evidence from different kinds of communities show that an essential mechanism for supporting cooperation is gossip and reputational threats can effectively promote trust (Greif 1993).

Giardini et al. (2014) developed a computational model in which they tested whether different reputation-based strategies may have an effect on cooperation rates in mixed populations. An essential element in the functioning of reputation is the action linked to it, although sometimes this action is implicit, for example, when that reputation is used to avoid cheaters. When partner selection is available, cheaters are avoided because of their reputations (Roberts 2008), but in indirect reciprocity models where cooperators cannot choose with whom to interact, players with bad image scores do not receive donations (Nowak and Sigmund 1998a).

In order to understand how the action linked to reputation might affect the overall cooperation levels, three different "reputation-based strategies" were defined, *as follows*:

- *Refuse* means that gossipers can refuse to contribute to the group when they know (on the basis of direct experience and gossip) that there is a majority of defectors in the group.
- *Compare* refers to the action of comparing between groups and actively looking for a better group.
- *Leader* is a refined form of partner choice in which group formation is delegated to a single agent, randomly selected to act as a leader and then allowed to choose its group mates. When the leader belongs to the population of "gossipers", it can use information received about others in order to select the best partners. A remarkable feature of this model is that information is privately transmitted among gossipers; therefore, it can become redundant and unreliable.

The results show that cooperation rates are higher when agents can compare their present situation and switch to a better one, i.e. when they can avoid free-riders, and this solution allows gossipers to get the highest scores in large groups of 25 agents. Moreover, the combination of punishment and gossip can make cooperation increase to its maximum in large groups, irrespective of the specific gossip strategy.

Group size can be a crucial factor, as showed by Suzuki and Akiyama (2005), who implemented a simulation model in which players in a PGG game can know other players' image score. In their work, cooperation can emerge in groups of four individuals, but increasing the size of groups inevitably leads to a decrease in the frequency of cooperation. The authors explain this result in terms of the limited observability of reputations in large communities with many individuals. In order to test whether this group size limitation also holds when agents are arranged on different networks, Vilone et al. (2016) compared two different network topologies, a small-world network and a bipartite graph. When reputation-based partner selection was available in a population distributed on a bipartite graph, full cooperation was reached after ten generations, also for larger groups of 20 individuals. This result has been replicated also with private gossip and errors in transmissions (Giardini and Vilone 2016) showing the importance of reputation in promoting informal social control and sustaining cooperation.

18.7 Conclusion and Future Work

Over the last two decades, there has been a significant increase in research on reputation and gossip. There is growing evidence that the presence of reputation can strongly promote cooperation and represents an effective way to maintain social control (Milinski 2016). Since reputation is a social coordination device emerging from the interplay of different information flowing in the social space, it could be difficult to test for emerging dynamics in a laboratory. This is especially true if we want to verify the difference in the usage of information with and without an explicit source, and we want to measure such a difference emerging from multiple interactions. Using artificial agents, i.e. computer programmes that behave according to some rules defined by the experimenter, we are able to investigate the complex interplay between the micro level of agents' motivations and the macro level of collective behaviours.

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 for partner selection in competitive settings like markets and for cheater isolation and punishment in cooperative settings like teamwork and grouping.

To exemplify both functionalities, we introduced two major simulation models of reputation in artificial societies. 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, neutralizing 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 Sim-Norm, 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. Socrate is an attempt to combine 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 useful to punish cheaters, but it also prevented the social network from collapse. We also discussed agent-based models of the evolution of cooperation in which gossip and punishment were compared as tools for social control, showing the importance of the former as an informal way of sanctioning non-cooperators.

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 do several things, as follows:

(a) Answer the question as to 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)
(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) and 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). 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). More recently, Hendrikx, Bubendorfer and Chard (2014) published a review of existing reputation systems, and the book by Bertino and Matei (2014) illustrated a project for the study of reputation in Wikipedia.

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 (2004) 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. 2004).

Further advanced issues for specialized 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. (1997). A distributed trust model. In *Proceedings of the 1997 Workshop on New Security Paradigms*. Langdale: ACM.

Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanisms. *The Quarterly Journal of Economics, 84*, 488–500.

Alexander, R. (1987). *The biology of moral systems*. New York: De Gruyter.
Bachmann, R. (1998). In T. Malsch (Ed.), *Kooperation, vertrauen und macht in systemen verteilter künstlicher intelligenz. Eine vorstudie zum verhältnis von soziologischer theorie und technischer modellierung* (Sozionik ed., pp. 197–234). Berlin: Sigma.
Barkow, J. H. (1996). Beneath new culture is old psychology: Gossip and social stratification. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind. Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
Beersma, B., & Van Kleef, G. A. (2011). How the grapevine keeps you in line: Gossip increases contributions to the group. *Social Psychological and Personality Science, 2*, 642–649.
Bertino, E., & Matei, S. A. (Eds.). (2014). *Roles, trust, and reputation in social media knowledge markets: theory and methods (computational social sciences)*. Berlin: Springer-Verlag.
Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge: Harvard University Press.
Bolton, G., Katok, E., & Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science, 50*(11), 1587–1602.
Cabral, L., & Hortaçsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics, 58*(1), 54–78.
Capraro, V., Giardini, F., Vilone, D., & Paolucci, M. (2016). Partner selection supported by opaque reputation promotes cooperative behavior. *Judgement and Decision Making, 11*(6), 589–600.
Carbo, J., Molina, J. M., & Davila, J. (2002). Comparing predictions of SPORAS vs. a fuzzy reputation agent system. In *3rd International Conference on Fuzzy Sets and Fuzzy Systems* (pp. 147–153). Switzerland: Interlaken.
Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior, 60*(1), 31–51.
Castelfranchi, C. (1998). Modelling social action for AI agents. *Artificial Intelligence, 103*, 157–182.
Castelfranchi, C., Conte, R., & Paolucci, M. (1998). Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation, 1*(3). http://jasss.soc.surrey.ac.uk/1/3/3.html
Conte, R. (1999). Social intelligence among autonomous agents. *Computational and Mathematical Organization Theory, 5*, 202–228.
Conte, R., & Castelfranchi, C. (1995). *Cognitive and social action*. London: London University College of London Press.
Conte, R., & Paolucci, M. (1999). Reproduction of normative agents: A simulation study. *Adaptive Behaviour*, special issue on Simulation Models of Social Agents, *7*, 301–322.
Conte, R., & Paolucci, M. (2002). *Reputation in artificial societies: Social beliefs for social order*. Norwell: Kluwer Academic.
Conte, R., & Paolucci, M. (2003). Social cognitive factors of unfair ratings in reputation reporting systems. In J. Liu, C. Liu, M. Klush, N. Zhong, & N. Cercone (Eds.), *Proceedings of the IEEE/WIC International Conference on Web Intelligence-WI 2003* (pp. 316–322). Halifax: IEEE Computer Press.
Conte, R., Castelfranchi, C., & Dignum, F. (1998). Autonomous norm-acceptance. In *Intelligent agents*. Berlin: Springer.
Cravens, K., Goad Oliver, E., & Ramamoorti, S. (2003). The reputation index: Measuring and managing corporate reputation. *European Management Journal, 21*(2), 201–212.
Dellarocas, C. N. (2003). *The digitalization of word-of-mouth: Promise and challenges of online feedback mechanisms* (MIT Sloan Working Paper 4296-03).
Di Tosto, G., Giardini, F., & Conte, R. (2010). Reputation and economic performance in industrial districts: Modelling social complexity through multi-agent systems. In K. Takadama, C. Cioffi-Revilla, & G. Deffuant (Eds.), *Simulating interacting agents and social phenomena* (pp. 165–176). Tokyo: Springer.
Diekmann, A., Jann, B., Przepiorka, W., & Wehrli, S. (2014). Reputation formation and the evolution of cooperation in anonymous online markets. *American Sociological Review, 79*(1), 65–85.

Dreber, A., Rand, D., Fudenberg, D., & Nowak, M. (2008). Winners don't punish. *Nature, 452*, 348–351.

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

Fehr, E., & Gachter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economics Perspectives, 14*, 159–181.

Fioretti, G. (2005). Agent based models of industrial clusters and districts. In *Contemporary issues in urban and regional economics*. New York: Nova Science.

Fombrun, C. (1996). *Reputation*, in Wiley Encyclopedia of Management.

Fombrun, C., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal, 33*(2), 233–258.

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*(1), 36–46.

Gardner, R., Ostrom, E., & Walker, J. (1990). The Nature of Common-Pool Resource Problems. *Rationality and Society, 2*, 335–358.

Giardini, F., & Conte, R. (2012). Gossip for social control in natural and artificial societies. *Simulation, 88*(1), 18–32.

Giardini, F., & Vilone, D. (2016). Evolution of gossip-based indirect reciprocity on a bipartite network. *Scientific Reports, 6*, 37931.

Giardini, F., Di Tosto, G., & Conte, R. (2008). A model for simulating reputation dynamics in industrial districts. *Simulation Modelling Practice and Theory, 16*, 231–241.

Giardini, F., Paolucci, M., Adamatti, D., & Conte, R. (2014). Group size and gossip strategies: An ABM tool for investigating reputation-based cooperation. In *MABS 2014: Multi-agent-based simulation XV* (pp. 104–118). Cham: Springer-Verlag.

Gluckman, M. (1963). Gossip and scandal. *Current Anthropology, 4*, 307–316.

Gray, E. R., & Balmer, J. M. T. (1998). Managing corporate image and corporate reputation. *Long Rage Planning, 31*(5), 695–702.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition. *American Economic Review, 83*(3), 525–548.

Hales, D. (2002). Group reputation supports beneficent norms. *Journal of Artificial Societies and Social Simulation*. *5*(4). http://jasss.soc.surrey.ac.uk/5/4/4.html

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

Hendrikx, F., Bubendorfer, K., & Chard, R. (2015). Reputation systems: a survey and taxonomy. *Journal of Parallel and Distributed Computing, 75*, 184–197.

Huynh, D., Jennings, N. R., & Shadbolt, N. R. (2004). Developing an integrated trust and reputation model for open multi-agent systems. In *Proceedings of the Workshop on Trust in Agent Societies (AAMAS-04)* (pp. 65–74). New York: ACM.

Jøsang, A., Ismail, R., & Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems, 43*(2), 618–644.

Koenig, S., Hudert, S., Eymann, T., & Paolucci, M. (2008). Towards reputation enhanced electronic negotiations for service oriented computing. In R. Falcone, S. K. Barber, J. Sabater-Mir, & M. P. Singh (Eds.), *Trust in agent societies—11th International Workshop, TRUST 2008*, Estoril, Portugal, May 12–13, 2008. Revised selected and invited papers. Berlin: Springer.

Kreps, D., & Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory, 27*, 253–279.

Marsh, S. (1992). Trust and reliance in multi-agent systems: A preliminary report. MAAMAW'92, 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Rome.

Marsh, S. (1994a). *Formalising trust as a computational concept*. Stirling: Department of Computing Science and Mathematics, University of Stirling.

Marsh, S. (1994b). Optimism and pessimism in trust. In *Proceedings of the Ibero-American Conference on Artificial Intelligence (IBERAMIA-94)*. Caracas: McGraw-Hill.

Melnik, M. I., & Alm, J. (2002). Does a seller's ecommerce reputation matter? Evidence from ebay auctions. *The Journal of Industrial Economics, 50*(3), 337–349.

Miceli, M., & Castelfranchi, C. (2000). The role of evaluation in cognition and social interaction. In K. Dautenhahn (Ed.), *Human cognition and social agent technology*. Amsterdam: Benjamins.

Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1687), 20150100.

Moukas, A., Zacharia, G., & Maes, P. (1999). Amalthaea and histos: multiagent systems for WWW sites and reputation recommendations. In M. Klusch (Ed.), *Intelligent information agents. Agent-based information discovery and management on the internet* (pp. 292–322). Berlin: Springer.

Muller, G., & Vercouter, L. (2008). *L.I.A.R. achieving social control in open and decentralised multi-agent systems* (Technical report). Saint-Étienne: École Nationale SupÉrieure des Mines de Saint-Étienne.

Nardin, L. G., Brandão, A. A., Sichman, J. S., & Vercouter, L. (2008) SOARI: A service oriented architecture to support agent reputation models interoperability. In R. Falcone, S. K. Barber, J. Sabater-Mir, & M. P. Singh (Eds.), *Trust in agent societies: 11th International Workshop, TRUST 2008, Estoril, Portugal, May 12–13, 2008. Revised selected and invited papers, no. 5396 in Lecture notes in artificial intelligence*. (pp. 292–307). Berlin: Springer. doi:https://doi.org/10.1007/978-3-540-92803-4_15.

Nardin, L. G., Brandão, A., Kira, E., & Sichman, J. S. (2014). Effects of reputation communication expressiveness in virtual societies. *Computational and Mathematical Organization Theory, 20*(2), 113–132.

Noë, A., & Hammerstein, P. (1994). Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology, 35*, 1–11.

Noon, M., & Delbridge, R. (1993). News from behind my hand: Gossip in organizations. *Organization Studies, 14*, 23–36.

Nowak, M. A., & Sigmund, K. (1998a). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573–577.

Nowak, M. A., & Sigmund, K. (1998b). The dynamics of indirect reciprocity. *Journal of Theoretical Biology, 194*, 561–574.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291–1298.

Ohtsuki, H., Iwasa, Y., & Nowak, M. A. (2009). Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature, 457*(7225), 79–82.

Paine, R. (1967). What is gossip about? An alternative hypothesis. *Man, 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. *Advances in Complex Systems, 3*(1-4), 39–51.

Paolucci, M., & Conte, R. (2009). Reputation: Social transmission for partner selection. In G. P. Trajkovski & S. G. Collins (Eds.), *Handbook of research on agent-based societies: Social and cultural interactions* (pp. 243–260). Hershey: IGI.

Paolucci, M., & Sichman, J. S. (2014). Reputation to understand society. *Computational and Mathematical Organization Theory, 20*(2), 211–217.

Paolucci, M., Eymann, T., Jager, W., Sabater-Mir, J., Conte, R., Marmo, S., et al. (2009). *Social knowledge for e-governance: Theory and technology of reputation*. Rome: ISTC-CNR.

Petroni, F., Querzoni, L., Beraldi, R., & Paolucci, M. (2016). LCBM: A fast and lightweight collaborative filtering algorithm for binary ratings. *Journal of Systems and Software, 117*, 583–594.

Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior, 29*, 172–178.

Pinyol, I., & Sabater-Mir, J. (2013). Computational trust and reputation models for open multi-agent systems: A review. *Artificial Intelligence Review, 40*(1), 1–25.

Pinyol, I., Paolucci, M., Sabater-Mir, J., & Conte, R. (2008). Beyond accuracy. Reputation for partner selection with lies and retaliation. In L. Antunes, M. Paolucci, & E. Norling (Eds.), *Multi-agent-based simulation VIII* (pp. 128–140). Berlin: Springer.

Porter, M. (1998). Clusters and the new economics of competition. *Harvard Business Review, 76*, 77–90.

Quattrociocchi, W., Paolucci, M., & Conte, R. (2008). Reputation and uncertainty reduction: Simulating partner selection. In *Proceedings of the International Workshop on Trust in Agent Societies* (pp. 308–325). Berlin: Springer.

Ramchurn, S. D., Huynh, D., & Jennings, N. R. (2004). Trust in multiagent systems. *The Knowledge Engineering Review, 19*(1), 1–25.

Rand, D. G., Arbesman, S., & Christakis, N. A. (2011). Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 19193–19198.

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. New security paradigms workshop, Lake Arrowhead, California, United States. In *Proceedings of the 1996 Workshop on New Security Paradigms* (pp. 18–25). New York: ACM.

Regan, K., & Cohen, R. (2005). Indirect reputation assessment for adaptive buying agents in electronic markets. In *Baseweb workshop at AI 2005 Conference*.

Resnick, P., & Zeckhauser, R. (2001). *Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system* (Working Paper for the NBER Workshop on Empirical Studies of Electronic Commerce).

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in internet transactions: empirical analysis of eBay's reputation system. In M. R. Baye (Ed.), *The economics of the internet and E-commerce*, *Volume 11 of Advances in Applied Microeconomics*. Amsterdam: Elsevier Science.

Ripperger, T. (1998). *Ökonomik des vertrauens: Analyse eines organisationsprinzips*. Tübingen: Mohr Siebeck.

Roberts, G. (1998). Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society of London B, 265*(1394), 427–431.

Roberts G. (2008). Evolution of direct and indirect reciprocity. Proceedings of the Royal Society B: Biological Sciences, *275*(1631), 173–179.

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. *European Mangement Journal, 22*(2), 201–210.

Saam, N., & Harrer, A. (1999). Simulating norms, social inequality, and functional change in artificial societies. *Journal of Artificial Societies and Social Simulation*, *2*(1). http://jasss.soc.surrey.ac.uk/2/1/2.html

Sabater, J. (2004). Evaluating the ReGreT system. *Applied Artificial Intelligence, 18*(9–10), 797–813.

Sabater, J., & Paolucci, M. (2007). On representation and aggregation of social evaluations in computational trust and reputation models. *International Journal of Approximate Reasoning, 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 & Multiagent Systems (AAMAS)*, Bologna, Italy. New York: ACM.

Sabater, J., & Sierra, C. (2004). Review on computational trust and reputation models. *Artificial Intelligence Review, 24*(1), 33–60.

Sabater, J., Paolucci, M., & Conte, R. (2006). Repage: REPutation and ImAGE among limited autonomous partners. *Journal of Artificial Societies and Social Simulation*, *9*(2). http://jasss.soc.surrey.ac.uk/9/2/3.html

Schillo, M., Funk, P., & Rovatsos, M. (2000). Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence, 14*, 825–848.

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

Sommerfeld, R. D., Krambeck, H., & Milinski, M. (2008). Multiple gossip statements and their effects on reputation and trustworthiness. *Proceedings of the Royal Society B: Biological Sciences, 275*, 2529–2536.

Squazzoni, F., & Boero, R. (2002). Economic performance, inter-firm relations and local institutional engineering in a computational prototype of industrial districts, *Journal of Artificial Societies and Social Simulation*, *5*(1). http://jasss.soc.surrey.ac.uk/5/1/1.html

Suzuki, S., & Akiyama, E. (2005). Reputation and the evolution of cooperation in sizable groups. *Proceedings of the Royal Society of London B: Biological Sciences, 272*(1570), 1373–1377.

Vilone, D., Giardini, F., & Paolucci, M. (2016). Exploring reputation-based cooperation: Reputation-based partner selection and network topology support the emergence of cooperation in groups. In F. Cecconi (Ed.), *New frontiers in the study of social phenomena* (pp. 101–114). Cham: Springer.

Wang, J., Suri, S., & Watts, D. J. (2012). Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 14363–14368.

Wilson, D. S., Wilczynski, C., Wells, A., & Weiser, L. (2000). Gossip and other aspects of language as group-level adaptations. In C. Heyes & L. Huber (Eds.), *The evolution of cognition*. Cambridge: MIT Press.

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific Reports, 6*, 23919.

Yamagishi, T., Matsuda, M., Yoshikai, N., Takahashi, H., & Usui, Y. (2004). Solving the lemons problem with reputation. An experimental study of online trading. In *eTrust: Forming relationships in the online world*. New York: Russell Sage Foundation.

Yu, B., & Singh, M. P. (2000). A social mechanism of reputation management in electronic communities. In *Proceedings of the 4th International Workshop on Cooperative Information Agents IV: The Future of Information Agents in Cyberspace*. Berlin: Springer-Verlag.

Zacharia, G. (1999). Trust management through reputation mechanisms. In C. Castelfranchi, R. Falcone, & B. S. Firozabadi (Eds.). *Proceedings of the Workshop on Deception, Fraud and Trust in Agent Societies, Seattle* (pp.163–167).

Zacharia, G., Moukas, A., & Maes, P. (1999). Collaborative Reputation Mechanisms in Electronic Marketplaces. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Wailea, Maui.