JASSS

[Jordi Sabater, Mario Paolucci and Rosaria Conte](#) (2006)

Repage: REPutation and ImAGE Among Limited Autonomous Partners

Abstract

This paper introduces Repage, a computational system that adopts a cognitive theory of reputation. We propose a fundamental difference between image and reputation, which 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'. 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, which we are in the course of developing for the public domain. In the paper, after a short presentation of the cognitive theory of reputation and its motivations, we describe the implementation of Repage.

Keywords:
Reputation, Agent Systems, Cognitive Design, Fuzzy Evaluation

Autonomy and Reputation in Social Agents

1.1

Reputation is a multi-purpose social and cognitive artefact, which probably co-evolved with human language and social organization ([Dunbar 1998](#)). Besides allowing for partner selection in exchange, reputation incentives cooperation and norm abiding and discourages defection and free-riding, handing out to nice guys a weapon for punishing transgressors by cooperating at a meta-level, i.e. at the level of information exchange ([Conte and Paolucci 2003](#)). The role of reputation as a partner selection mechanism started to be appreciated in the early eighties ([Kreps and Wilson 1982](#)). However, little understanding of its cognitive underpinnings was achieved at that stage. Evolutionary game theorists ignored the difference between image (i.e. own believed evaluation of a target, see [Nowak and Sigmund 1998](#)) and reputation (i.e. nested evaluation or meta-evaluation: a belief about how a given target is commonly said to be evaluated). Consequently, the decision to report on reputation to others, whether one shares it or not, was ignored. Hence, the efficacy of preventive social knowledge was not fully appreciated and, what is worse, the role of reputation in updating existing social evaluations was overlooked. The distinction between image and reputation, i.e. between beliefs and meta-beliefs, brings about the necessity for a cognitive approach to the subject matter. This starts now to be perceived as fundamental in the study of reputation (see for example [Grunig and Hung 2002](#), which however is focused on the mental *effects* of reputation, rather than on the cognitive nature of the phenomenon). Until very recently, the cognitive nature of reputation was substantially ignored. This has caused a misunderstanding of the effective role of reputation in many real-life domains and the related scientific fields.

1.2

A special field of application, which is becoming ever more important, is the effect of reputation in virtual and agent-mediated markets. Classic systems like eBay are known to exhibit a characteristic bias to exceedingly positive evaluations ([Resnick and Zeckhauser 2002](#), [Bolton et al. 2002](#)), suggesting that factual cooperation among users at the information level may lead to a 'courtesy equilibrium', which is actually neither fair nor efficient. The design of those systems draws attention on the cognitive side of the phenomenon ([Conte and Paolucci 2002](#)).

1.3

In this paper, we describe Repage, a computational system for partner selection. Although we will show here an application of Repage to a competitive setting (marketplace), in principle it can also be used in cooperative contexts (organizations). As we shall see, based on a model of REPutation (beliefs about shared voice on a given target), imAGE (own evaluations), and their interplay, Repage provides evaluations on 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. If agents transmit only believed image, the

circulation of social knowledge is bound to stop soon.

1.4

But in order to preserve their autonomy, agents need to *decide* whether to share or not others' evaluations of a given target. If agents would automatically accept and transmit as their own reported on evaluations, they would be no more autonomous. Hence, they must

- form both evaluations (image) and meta-evaluations (reputation), keeping distinct the representation of own and others' evaluations, before
- decide whether or not to integrate reputation with their own image of a target.

1.5

In addition, in order to exchange information about reputation, agents ought to participate in circulating it whether they believe it or not (gossip); but to preserve their autonomy, they must decide how, when and about whom to gossip. In sum, the distinction between image and reputation is derived from the paradox of sociality, i.e. the trade-off between agents' autonomy and their need to adapt to social environment, being open to social influence. At the same time, thanks to such a distinction, agents are provided with the means for coping with the paradox in question.

1.6

However, the interplay between image and reputation leads to both uncertainty and inconsistency. Inconsistencies do not necessarily lead to a state of cognitive dissonance, nor do they always urge the system to find a solution. For example, as we shall see later on in the paper, an inconsistency between own image of a given target and its reputation creates no serious problem to the Repage system, which can keep track of both while giving more weight to experience than to others' communication. It is true that agents do sometimes prefer to stick to their own evaluations, even when disconfirmed by others or by events. However this is always a matter of decision. Agents might decide to ignore disconfirming evidence and the question of course is when and why they do so. What we need is a model of such a decision. Moreover, agents might stick to their first impression, but transmit others' evaluations ("I like that guy, but others say he is a cheater"). Even without mistakes or noise, reputation is spread even if it is not believed. It soon diverges from evaluations, although interacting with them.

1.7

Actually, a contradiction between own evaluations is sometimes possible: I may get a good impression from a given experience with the target, which may be dismantled next time. Or, my direct experience may be confirmed in further interaction, but at the same time it may be challenged by the image I believe others, whom I trust a lot, have formed about the same target. In both cases, I find myself in a rather awkward condition, especially when that target is one of the few, if not the only, available partners for a necessary transaction. What will I do in such a condition? Will I go ahead and sign a contract — maybe a low-cost one, just to acquire a new piece of direct evidence — or will I check the reliability of my informants? Suppose the latter alternative is chosen on the grounds of a cost to benefit rule. What does it mean to check others' reliability? If their image of the target is better than own, and if one should discard direct but costly experience, what else should one do?

1.8

The picture is rather complex, and the number of possibilities is bound to increase at any step, making computationally heavy the application of rule-based reasoning. However, this complexity mimics the one that is present in the reality of cognitive elaboration; its oversimplifications, like the evaluation mechanisms set up in eBay-like electronic markets, are a poor substitute of the real thing. In this paper, we want to set the scene for a system that does not hide the complexity of the cognitive constructs used for reputation and for their relationships, and to present our implementation of such a system, showing that it possesses the desired characteristics — meaningful construction and manipulation of image and reputation, whose interplay is necessary for agents forming and updating social evaluations.

1.9

In the following, we will offer a brief presentation of the theory of reference on reputation, followed by a short discussion on the proposed fuzzy representation of evaluations. The Repage system in its current implementation will then be presented and discussed. Some situations, aimed at illustrating how Repage operates, will be described and compared with a similar system, ReGreT.

🌍 Value Added of Reputation: Theory of Reference

2.1

The social cognitive perspective on reputation presented in this paper aims to model both the cognitive properties and the social aspects of reputation, including its transmission. In order to model both, it is necessary to understand the interrelationships between two different types of social evaluation, i.e. *image* and *reputation*. We present here, in short, the basic components of a theory of reputation built as an extension of Conte and Paolucci ([2002](#)).

2.2

Image and reputation are distinct objects. Both are social evaluations. They concern other agents' (targets) attitudes toward socially desirable behaviour, and may be shared by a multitude of agents. But whereas image consists of a set of evaluative beliefs about the characteristics of a target, reputation concerns the voice that is circulating on the same target — and in general, there is no motivation to think that they will coincide. Both notions concern the evaluation of a given object, more specifically of a social agent, which may be either individual or supra individual, and in the latter case, either a group or a collective.

2.3

More in detail, image is an evaluative belief ([Miceli and Castelfranchi 2000](#)); it tells that the target is "good" or "bad"

with respect to a norm, a standard, or a skill. (Social) evaluations may concern physical, mental and social properties of targets. In particular, agents may evaluate a target as for its capacity as well as its willingness to achieve a shared goal. The interest/goal with regard to which a target is evaluated may also be a distributed or collective advantage.

2.4

At the meta-level, *reputation* is a belief about the existence of a *communicated* evaluation, more specifically about a related but somehow impersonal evaluation of the target. This bears several important consequences. First, to accept a meta-belief does not imply to accept the contained belief. Consequently, to assume that a target *t* is assigned a given reputation implies only to assume that *t* is reputed to be "good" or "bad", i.e., that this evaluation circulates, but it does not imply to share the evaluation.

2.5

To be more precise, at the next level we distinguish between *shared evaluation* and *shared voice*. A given agent may have a belief about others' evaluations of a target. When these evaluations converge, we say that the evaluation of the target is shared among a given set of agents. A shared evaluation is essentially a special case of an image. Even if it does not coincide with one's own image of a target, a shared evaluation is likely to be accepted by an individual agent, especially if those sharing it enjoy a good evaluation (in the role of information providers) in the latter's eyes. Note that in a shared evaluation all individuals constituting the set of sharing agents are precisely identified.

2.6

A shared voice is a similar object, whose content is removed another level to build a belief about others' beliefs on the existence of a voice that evaluates a target. In other words, the agent holding the shared voice believes that a precisely identified set of agents, if asked, will consistently report on the existence of a voice. While this is very nearly what we ask for a definition of reputation, there is still another level of abstraction to be added: the set of agents. With shared voice too, all individuals constituting the set of sharing agents are precisely identified.

2.7

Reputation in the full sense is built on all these ingredients by generalization and loss of reference; when an agent has a reputation in the proper sense, it means that a corresponding evaluation circulates in a group — meaning that most members of the group will agree that such a voice exists without specifying a precise set of referents. Compared to a shared voice, the referents set is now lost in its precise identification and substituted by a less precise group attribution. In comparison to shared evaluation — and more generally to image — reputation does not take a stand on what is true but just on what is told.

2.8

To understand better the difference between image and reputation, also the mental decisions based upon them must be analyzed. They consist of three decisions:

- Epistemic: accept the beliefs that form a given image or acknowledge a given reputation.
- Pragmatic-Strategic: use these beliefs in order to decide whether and how to interact with the target.
- Memetic: transmit these beliefs to others.

2.9

This difference is not inconsequential: to spread news about someone's reputation (shared voice) does not bind the speaker to commit herself to the truth value of the evaluation conveyed but only to the existence of rumours about it. Therefore, unlike ordinary sincere communication, only the acceptance of a meta-belief is required in communication about reputation. And unlike ordinary deception, communication about reputation implies:

- no personal commitment of the speaker with regard to the main content of the information delivered. If the speaker reports on target *t*'s bad reputation, he is by no means stating that *t* deserved it; and
- no responsibility with regard to the credibility of (the source of) information ("I was told that *t* is a bad guy"). Two points ought to be considered here. First, the source of the meta-belief is implicit ("I was told"). Secondly, the set of agents to whom the belief *t* is attributed is none defined ("*t* is ill/well reputed").

Of course, this does not mean that communication about reputation is always sincere. Quite on the contrary, one can and does often deceive about others' reputation. However, to be effective, liars neither commit to the truth of the information transmitted nor take responsibility with regard to its consequences. If one wants to deceive another about somebody's reputation, one should report it as a rumour independent of or even despite one's own beliefs! As a consequence of this analysis, we can see how, unlike other (social) beliefs, reputation may spread in a population even if the majority does not believe it to be deserved. Meta-beliefs can spread without first-level beliefs spreading.

Current Systems

2.10

Applications of reputation abound in two sub-fields of information technologies, i.e. computerized interaction (with a special reference to electronic marketplaces) and agent-mediated interaction.

Online Reputation Reporting Systems

2.11

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). Indeed, some authors (Cabral and Hortaçsu 2004) do argue that courtesy is a long-run equilibrium in eBay. As they formally prove, initial negative feedbacks trigger a decline in sale price that drives targets out of the market. Instead, better sellers have more to gain from 'buying a reputation' by building up a record of

favourable feedback through purchases rather than sales. Two effects follow: those who suffer a bad reputation stay out, at least until they decide to change identity; those who stay in can but enjoy a good reputation: after a good start, they will hardly get negative feedbacks, and even if they do, these will not get to the point of spoiling their good name. Of course, under such conditions, even good sellers may have an incentive to sell lemons.

2.12

Intuitively, the courtesy equilibrium reduces the deterrent efficacy of reputation. If a reputation system is meant to reduce frauds and improve the quality of products, it ought 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. Instead, we believe that, rather than suggesting remedies or recipes based on local arguments and fragmented models, a general theory of how reputation and its transmission works ought to be worked out and on top of such a theory, different systems for different objectives ought then to be constructed.

MAS Applications

2.13

Models of reputation for multi agent systems applications ([Yu and Singh 2002](#); [Carbo et al. 2002](#); [Sabater and Sierra 2002](#); [Schillo et al. 2000](#); [Huynh et al. 2004](#)) clearly present interesting new ideas and advances over conventional online reputation systems, and more generally over the notion of global reputation, or centrally controlled image. Indeed, models of trust and reputation abound in this field (for a couple of exhaustive reviews, see [Ramchurn et al. 2004a](#); [Sabater and Sierra 2004](#)).

2.14

As can be observed, 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 one another. Reliability of agents implements 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 users' interactions. In a sense, implemented agent systems for agent-mediated interaction represent both parallel and nested sub-communities.

2.15

As a consequence, solutions applied to the problems encountered in this environment are validated more severely, against both external and internal criteria. Second, 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. Third, and moreover, solutions may be not (only) implemented between the agents, but (also) within the agents, what greatly expands the space for modelling.

2.16

So far, however, these potentialities have not been fully exploited. Rather than research-based systems for reputation, models have been aimed to ameliorate existing tools, implemented for computerized markets. 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 agent and processes are needed for which result: better efficiency, encourage equity (and hence users' trust), discourage either positive or negative discrimination (or both), foster collaboration at the information level or at the object level (or at both), etc.

2.17

The solutions proposed are interesting but insufficient attempts to meet the problems left open by online systems. Personalization and group reputation solutions are useful to the extent that they apply to small size groups thereby under-exploiting the reputation mechanism. Collective filtering and sanctioning should be based upon transmission, so that neither evaluations collapse on community or reliable reputation nor the other way around. More extensive investigation of the reciprocal effect of these two notions (see the previous section) is needed, in order to establish what leads to undesirable non-trivial phenomena like coalitions and discrimination. In a few words, what is strongly needed is a theory-driven design of reputation systems,

- based upon a conceptual analysis of the differences and analogies among notions concerning social evaluation.
- implying a non-atomized view of interaction and aimed to exploit inter-agent processes.

2.18

As is often the case, to understand how things work in real societies is necessary for improving the performance of technologies. Reputation is an old remedy for a problem (find out the bad guys) that is affecting human societies since they started to enlarge. We set out to understand how it works in natural societies, and to design reputation systems once we know better what can be expected of them, to what extent and under which conditions. The results of such an endeavour should include a set of recommendations that could help in improving existing system like eBay. However, this is an objective that cannot be reached but stepwise — the first step being that of building a working system, capable of manipulating separately image and reputation. In this work, we aim to illustrate our proposal — what it is and how it works — for the implementation of a (rather complex) system with these characteristics.

🌍 Fuzzy evaluations and their composition

3.1

In this section, we define our proposal for the representation of elementary information in Repage. We need a representation of a social evaluation (Miceli and Castelfranchi 2000) that will allow both communications and personal evaluations to be represented.

3.2

In most real social situations, there is no way to make social evaluations precise; the main exception being economic science, blessed by the invention of money. It is not by accident that economics has been the first social science to be mathematically and formally studied. To the contrary, most of human social skills are based on imprecise and incomplete data, made even vaguer by the tendency to misrepresent frequencies. However, it is hard to deny that the performance of human society is impressive, and we are still far from understanding it in detail. The purpose of Repage is that of providing advances both in designing and in understanding how a cognitive approach — that takes into account the subtleties inspired by the analysis of human cognitive artefacts — could be more advantageous than a purely rational one, based on game theory and simple (usually numeric) representations of reputation.

3.3

To decide how to handle social evaluations on the basis of the theory presented above, we have to consider them in three different aspects.

3.4

The first aspect is the *type* of evaluation — as discussed above, personal experience, image, third party image, shared voice and all the other cognitive constructs have different functional properties that call for a clear and sharp distinction. In our model, different types of information go through different paths in the cognitive network that represent the memory. This is a sharp distinction; interaction between different types of information is regulated externally. The assumption here is that there will be no intrinsic noise or lack of precision in distinguishing between the types — for example, the agents will not confuse the results of direct experience with related information, nor will they confuse reputation information — what other agents say that "is generally said" — with information having a well defined source — image or third party image.

3.5

The second aspect is the *subject* (or *role*) evaluation concerns. Are we considering our target as a seller or as an informant? In our system, surely we want to keep different aspects separated, at least to some extent. In all of the examples that will be considered, we separated evaluation of exchange performance (target as seller) from evaluation of communication (target as informer, separately for image and for reputation). This distinction also shows interesting functional properties: changes in the evaluation of somebody as an informer will reverberate on all evaluations the agent gets from that source, adjusting its strength accordingly. Again, we treat this distinction sharply — a piece of information can regard an agent as an informer or as a seller, but not both.

3.6

The third aspect is the *content* of the evaluation: is John good or very good? 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; one is not told that people is saying around that Jane is 0.234 good. While we always identify precisely the type of information communicated and the role discussed, we want to leave some space in the evaluation itself to capture the lack of precision coming (a) from vague utterances, i.e. "I believe that agent *t* is good, I mean, very good — good, that is", and (b) from noise in the communication or in the recollection from memory. For these reasons, we decided to model the actual value of an evaluation with a fuzzy set, represented by a tuple of positive real values that sum to one (Zadeh 1965). These values express the membership of the evaluation to a rating scale. For this version of the model, we tried with five different levels, ranging from very bad to very good. Moreover, we add to the number a value indicating the belief strength in the evaluation. By this means, we are able to represent both *x*'s strong beliefs that *y* is moderately good and, on the opposite, *x*'s mild confidence in the general assumption that *y* is an angel.

3.7

The addition of fuzziness in the evaluation requires several delicate decisions to be taken: having decided what kind of fuzzy set we want to use, we need to define carefully how to operate on them by weighting, aggregating, and comparing. For all these issues we propose solutions (mostly standard ones) from the literature, presented briefly in the following subsection.

Use of fuzzy sets in Repage

3.8

We use a tuple of five numbers, which sum to one, to represent the membership of our evaluation to the following rating scale: very bad, bad, neutral, good, very good. In mathematical terms, we have a tuple of positive numbers $\{w_1, w_2, ..., w_5\}$, whose sum is one, where w_1 corresponds to very bad and w_5 to very good. In addition, we have a single value indicating the strength of belief in the evaluation, an unbounded positive number *s*. In the following, we will sometimes express this evaluation as the 6-ple $\{w_1, w_2, ..., w_5, s\}$. In fig. 1, we show some examples.
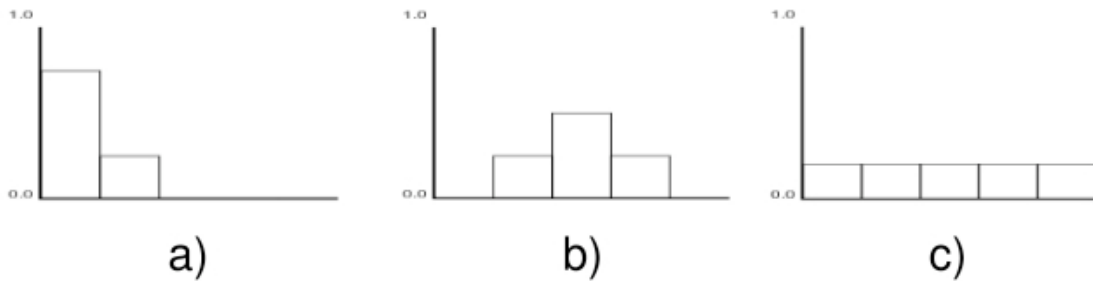
Figure 1. Examples of fuzzy evaluations. In a) we have {0.75, 0.25, 0, 0, 0}, a mostly bad (75% very bad, 25% bad) evaluation; b) is an average evaluation, neither good nor bad. In c), we show the situation of maximum uncertainty; aggregating another number with this one leaves the first unchanged.

3.9

This representation differs from the one chosen in Carbo et al. (2003), where a fuzzy number for reputation is represented by a function over the possible values in the [0,100] interval, an approach that looks more probabilistic than fuzzy in the proper sense. Moreover, the problem of fuzzy aggregation is resolved by taking the weighted mean between the new contribution and the previous value of reputation. In the following, we discuss the problems associated with this aggregation method. Fuzzy concepts are also employed in Ramchurn et al. (2004b), where fuzzy sets are used to relate confidence levels with the expected values for the issues in a contract, or in Falcone et al. (2003) where fuzzy maps are employed to describe trust. However, none of these systems show an elaboration on the different functional roles of image with respect to reputation; the cognitive elaboration of these artefacts is still minimal.

3.10

The proposed definition of our fuzzy evaluation is quite consequential, even if it contains some elements of arbitrariness. The key structure of our model, anyway, is not this representation choice, but the network of relationships between the cognitive constructs. For this reason, we believe that modifications to the proposed fuzzy representation will not substantially affect the workings of the model — for example, the modification of the scale of values should not influence much the results.

3.11

The decision to use fuzzy representations presented us with the problem of how the basic algebraic operations work on them (Yager 2004a); for all the basic operations, several choices exist. Let us review the three main operations needed for Repage.

Aggregation of evaluations

3.12

The main operation required in Repage is a weighted aggregation of fuzzy evaluations. The aggregation will be needed, for instance, when the agent creates an image or a reputation on the basis of several contributing beliefs. We consider the contributing beliefs to be already evaluated, for example by credibility of the source; we are interested here in the aggregation operation only for $j:1..n$ fuzzy evaluations whose weights we design by w_i^j, where the lower index i refers to the different weights of the same fuzzy evaluation, and the higher one j is used to distinguish the evaluations to aggregate. A detailed report of problems that can arise by not considering the details of this choice can be found in Yager (2004b), which will be used as our reference in the following discussion. The first natural choice, that is, the mean value for each element of the scale, also used in Carbo et al. (2003), is shown to have several unpleasant properties.

3.13

Let us consider the problem of identity, as a fuzzy set that, if aggregated with any other fuzzy sets, will leave the latter unchanged. In our representation, there is a natural candidate for identity, and that is the "flat number" $w_i = 1/k$ for all i, where k is the number of weights composing the fuzzy set (5 in our case); this number expresses total indifference to the levels and should leave other numbers under aggregation unchanged. To see this, consider the case in which a "very good" value is aggregated with an uncertain value: aggregation does not lead to modify the former value. Of course, aggregation by pure mean does not respect the natural value for identity.

3.14

Moreover, the aggregation function must be associative and commutative — the result must not depend by the order in which evaluations are aggregated and must be easy to calculate. Another important property is stability: an aggregation of many values should not "jump" by the addition of some special one, but should change more or less in a continuous way.

3.15

Several other aggregation functions are listed in Yager (2004b); all of them are based on the multiplication of values, that is

$$w_i = \frac{\prod_j w_i^j}{\sum_i \prod_j w_i^j}$$

that is also our choice for use in Repage. This function shows several good properties: besides respecting the identity,

it is commutative and associative. However, it loses sense when denominator — the sum of the products of weights — is zero; notice that, for this to happen, it is sufficient to have at least one zero in all evaluation levels for some of the values to be aggregated, which, for large aggregations, is quite easily the case. Yager (2004b) proposes several solutions, but in Repage we get rid of this problem as a side effect of the strength of beliefs. To take strength into account, we apply the standard procedure of moving the fuzzy set to be aggregated towards the identity by a quantity determined by the strength (rescaled to the [0, 1] interval by the use of the *arctan* function). After rescaling, no value can be zero if not in the case of infinite strength, thus avoiding the problem of incompatible values.

Calculating the strength of beliefs

3.16

As seen in the previous paragraph, augmenting the fuzzy representation with the strength of beliefs avoids what can be seen as the main problem with aggregation, i.e. incompatible values.

3.17

But how do we create or maintain beliefs' strength? We decided to follow the simplest strategies for what regards the strength of the basic elements, such as attributing to a direct experience belief strength proportional to the investment needed, that is, in a market case, the price of the transaction. The only case that needs to be discussed here occurs when the strength of a belief formed from a communicated evaluation must be weighted for the agent's opinion of its informer.

3.18

If we represent agents' evaluations as fuzzy set, in this case we will need to find out what we mean by the fact that our current opinion about our friend, the source, as an informer is {0, 0.2, 0.2, 0.6, 0; 2} — that is, we believe with strength 2 that our friend is one fifth bad, one fifth neutral, and three fifths good. What does this mean when we want to decide whether to form a belief from our friend's sentence saying that target is {0, 0, 0, 0.5, 0.5} as a seller? We do not want to change the content of the communication — indeed we want to maintain the fuzzy evaluation as it is, scaling only its strength. To rescale strength, we need to turn our opinion of the source into a scalar number. We decided to calculate a weighted sum of the evaluation levels values: the strength with which we accept the information is scaled by the sum of $(i-1)\,w_i$ on all levels, divided by the number of levels (5 in our case).

Comparing two evaluations

3.19

In some cases, for example when comparing the messages from two different agents, we need to compare two fuzzy values. Unlike real numbers, there is a large arbitrariness in the metrics that can be used to calculate their difference. Since we will always compare two values per time, we first calculate the difference in absolute value between the fuzzy values level by level. What we obtain (not a fuzzy set, since it is not normalized) is then used to calculate a sort of momentum, i.e. the sum of the products of the weights by their distance from the centre of mass. In formulas, if the values are w_i^1 and w_i^2 and the difference of the weights is $d_i=|\,w_i^1-w_i^2|$, the centre of mass and the distance are calculated as the following:

$$c = \sum_i d_i(i-1)/4 / \sum_i d_i; \ \ dist = \sum_i d_i \left| c - \tfrac{i-1}{4} \right|$$

In words, the higher the momentum, the wider the difference between the values compared.

🌍 Repage: Architecture and Implementation

4.1

In this section we describe the architecture of the Repage model as well as how it is integrated with the other elements that compose a typical deliberative agent. An implementation of Repage has been developed by the authors in Java; the source code has been published as a Sourceforge project. The code used to produce all the examples in this paper is available by anonymous cvs at http://cvs.sourceforge.net/viewcvs.py/repage/papers/JASSS2005.

The architecture

4.2

The *Repage* architecture, as shown in fig. 2, is composed of three main elements: a memory, a set of components called *detectors* and the *analyzer*.
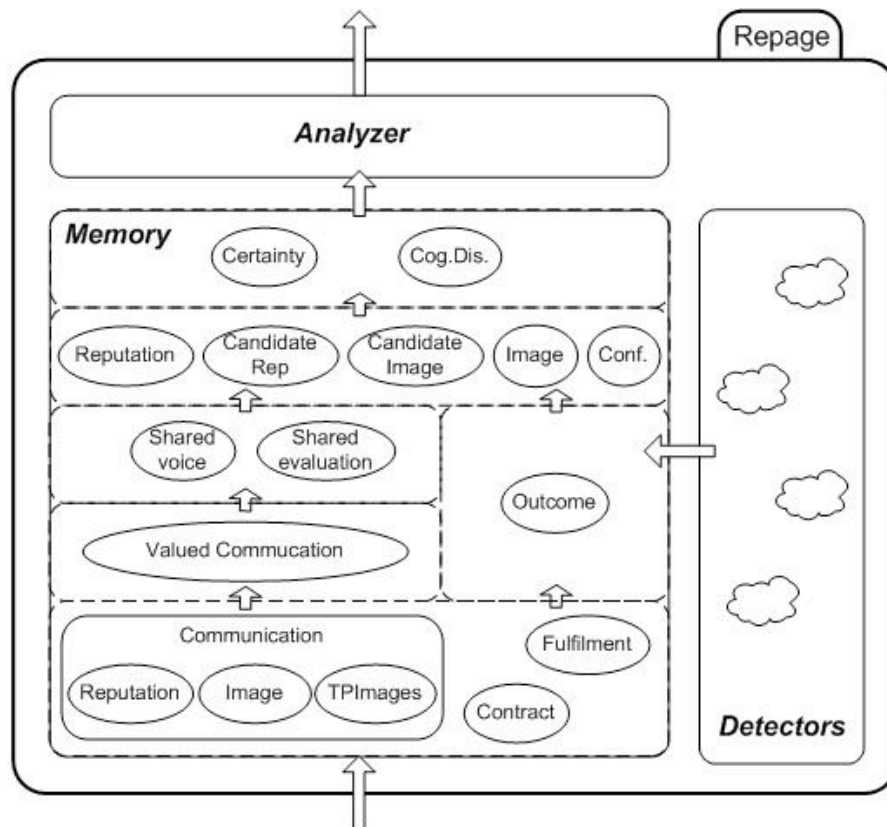
Figure 2. Repage architecture

Memory and the detectors

4.3

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. 3). 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.

4.4

To mirror their dependence connections, in the Repage memory predicates are conceptually organized in different levels and inter–connected. Each predicate reference is wrapped up by a component that adds connection capabilities to it. This approach allows the predicates in the main memory to be maintained clean, without conditioning their use to other modules of the agent.

4.5

Predicates contain a fuzzy evaluation, belonging to one of the main types (image, reputation, shared voice, shared evaluation), or to one of the types used for their calculation; these include valued info, evaluation related from informers, and outcomes. Special–purpose predicates, dependent on the application domain (for example, a contract not yet fulfilled), exist in the lower layer; they do not necessarily contain an evaluation. Each predicate (except the special purpose ones) has a role and a target; for example, an image of an agent (target) as informer (role).

4.6

Finally, each predicate has a strength value associated to it. This value is a function of (i) the strength of its antecedents and of (ii) some special characteristics intrinsic to that type of predicate. For instance, the strength of an Image is a function of the strengths of the antecedents (outcomes, information from third party agents and their image or reputation as informers) but also of the number of these antecedents. Taking into account the number of antecedents, however, has no sense in the case of an outcome because there are always two antecedents (the contract and its fulfilment).

4.7

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.

4.8

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 when these are no more useful and, more important, for creating the network of dependencies among the predicates.

4.9

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 the situation to all the detectors 'interested' in that type of predicate. This starts a cascade process where several detectors are activated

one after the other. At the same time, the dependency network ensures the correct actualization of the predicate values according to the new additions and subtractions.

4.10

Bias is another factor that influences the strength of predicates. Biases are rules that give more or less relevance to certain aspects of the situation. They are based on either socio-psychological studies (like for instance the stronger weight to negative rather than to positive information, see Skowronski and Carlston 1989) or common sense (for example, the stronger weight of direct than indirect experience). Two remarks are needed here. First, the existence of such a bias should not be intended to diminish the role of reputation since agents that are likely to weigh direct more than indirect experience in image formation, might be biased in the opposite sense when spreading reputation. Hence, the latter might have a stronger influence on agents that had no chance to meet with the target. Second, biases might be dropped or modified; although they form the initial knowledge of the agent, they are not static rules. What is intuitive under given conditions may turn out to be unwarranted in other. If the experience of the agent suggests it, a bias rule can be modified to fit with new environmental conditions. For example, direct experience is more reliable in domains characterized by none or poor expertise. In special domains, the opposite bias seems instead much more reasonable.

4.11

At the first level, predicates are not yet evaluated by Repage. They are the starting point of the inference task that will be performed by the detectors. At this level, there are three types of information:

- Contracts: In this context, a contract is not necessarily a formal contract. It can be just an agreement between two agents.
- Fulfilments, i.e. the results of contracts. For instance, in an e-commerce environment, these are the real characteristics of a product that the agent bought several days before.
- Informers' communication: information about other members of the community coming from third party agents (informers). The transmitted information can be related to three different aspects: the image that the informer has of the target, the image that according to the informer other agents have of the target (third party image) and finally the reputation of the target, which will contribute to the building of a shared voice.

4.12

Two detectors work at this level. One can infer a new outcome from a contract and its fulfilment; the other, given a certain communication, generates valued information. These two new predicates form the second level.

4.13

We define the outcome of a transaction between two agents as:

- An initial contract to take a particular course of action and the actual result of the actions taken.
- An initial contract to fix the terms and conditions of a transaction and the actual values of the transaction terms.

4.14

An outcome is not just the tuple contract-fulfilment, it is also the evaluation of this tuple considering how the contract was fulfilled. A piece of valued information is a communicated information once it has been evaluated according to the reliability of the informant. To perform this evaluation the Repage model uses the image of the informer, weighting the new information as for 3.18.

4.15

This led us to the next conceptual level. In this level we find two predicates: shared voice and shared evaluation. A shared voice is the main element to build a reputation. It is built from communicated reputation from third party agents. On the other hand we have the shared evaluation that is built from communicated images. With outcomes, shared evaluations are the main elements upon which to build up images.

4.16

In the fourth level there are five types of predicates: *Candidate Image*, *Candidate Reputation*, *Image*, *Reputation* and *Confirmation*. As the names indicate, candidate images and candidate reputations have no sufficient support yet, to become real images and reputations (either because elements contributing to them are insufficient, or because information is inconsistent). There is a specialized detector for each type. Once a candidate image or a candidate reputation reaches a certain level of strength it becomes a full image/reputation. The idea behind the last predicate, Confirmation, is that it mirrors how good previous information was. A communication, the truth-value of which is known to the recipient, feedbacks on the image/candidate image of the information sender as an informer. The Confirmation is similar to an outcome where the contract is the communication provided by the informer and the fulfilment is the image the agent has about the target. What differs from an outcome is the nature of the information. Both pieces of information (the image and the information provided by the informer) are far from objective. Therefore, the weight of a Confirmation usually is lower than that of an outcome.

4.17

In the last level, we find the last two predicates: *cognitive dissonance* and *certainty*. A cognitive dissonance is a contradiction between two pieces of information that are relevant for the individual and refer to the same target; it generates instability in the mind of the individual. Depending on how strong and relevant the cognitive dissonance is, the individual is pushed to take special actions for solving it. Although these actions are context dependent, they are always oriented to confirm the grounds of the elements that are causing the cognitive dissonance. On the other side, a certainty predicate implies a full reliance on what the certainty asserts.

4.18

It should be evident now that in this model reputation is not equal to uncertain image: the difference between image

and reputation is not only nor primarily a matter of certainty: an agent may have a strongly evaluated image of a target, and know that she enjoys from a completely different reputation, and may take decisions on the grounds of both.

The Analyzer

4.19

The main task of the analyzer is to propose actions that (i) can improve the accuracy of the predicates in the Repage memory and (ii) can solve cognitive dissonances trying to produce a situation of certainty. The analyzer can propose one or more suggestions to the Planner, letting it decide on its own criteria whether to execute them or not.

Repage and its environment

4.20

Repage has been designed having in mind an easy integration with the other elements that compose a deliberative agent. At the same time, this integration is one of the key points of the model. Repage is not only a passive module the agent can query to know about the image and reputation of another agent. The aim of Repage is also to provide the agent (in this case, the planner of the agent) with a set of possibilities that can be followed to improve the reliability of the provided information.

4.21

You can always get a result from Repage. However if you feed it with the right information, it will be more accurate. The analyzer is responsible for making the right proposals to improve the reliability of the relevant topics.

4.22

Section A in fig. 3 shows how Repage is integrated with the other elements of an agent (for the sake of brevity we show only those elements that interact with Repage).
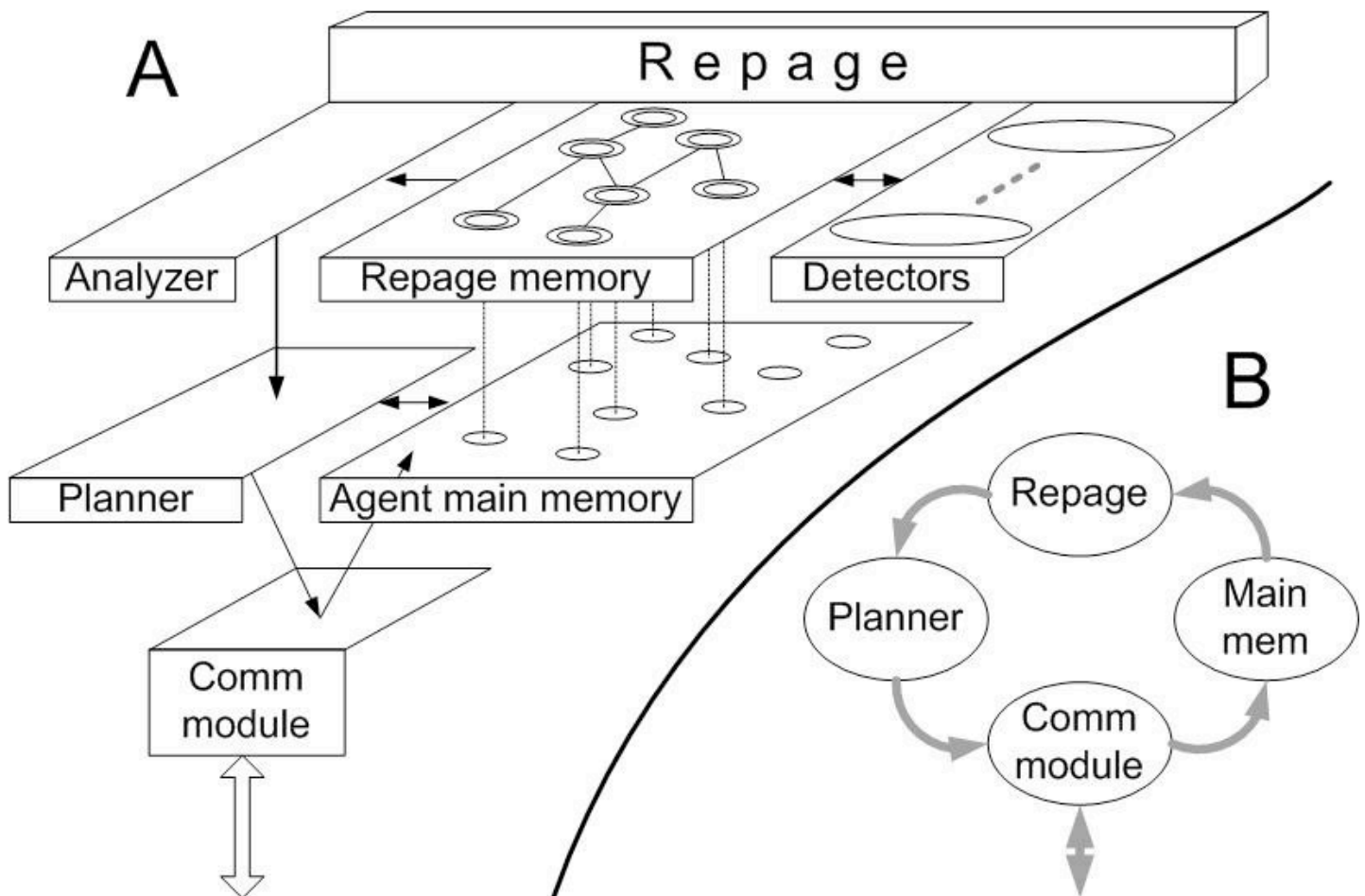
Figure 3. Repage and its environment

4.23

The Communication module is the module that connects the agent with the rest of the world. Here is where the fulfilment of the contracts and the results of communication arrive first. After a possible process of filtering/transformation, the predicates are added to the main memory. This addition has an immediate effect on the Repage memory that mirrors it.

4.24

The Repage memory is composed by a set of references to the predicates in the main memory of the agent that are relevant to deal with Images and Reputations. The actions of the detectors over the Repage memory imply the 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 in the main memory, the dependence network is present only in the Repage memory.

4.25

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. 3, section B.

Repage in action

5.1

In this section we will present two situations to illustrate some of the main points in the behaviour of Repage. To produce each situation we have used the current Repage implementation in Java (the code can be retrieved at http://cvs.sourceforge.net/viewcvs.py/repage/papers/JASSS2005). For each situation we will compare the operation of Repage with that of ReGreT (Sabater 2003), another system addressing reputation. Results from this comparison can easily be extended to other models using aggregations of direct and third party information.

5.2

The general scenario used to illustrate these situations is the following: *ag-0* is a buyer, and is endowed with a module processing reputation. She knows that *ag-T* sells what she needs but knows nothing about the quality of *ag-T* (the "Target" of the evaluations) as a seller. Therefore, it 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. *Informants* (that is, agents potentially able to provide information about the target and about other informants) will be indicated by $ag\text{-}I_x$ (*x*:1…n).

The ReGreT model

5.3

To understand the comparison with the ReGreT model we need to present, at least in its basics, how the ReGreT model works. For a detailed description refer to Sabater (2003).

5.4

In ReGreT there is no difference between evaluation and meta–evaluation, but only between what is called Direct Trust (calculated from direct experiences) and Reputation (calculated from the data coming from informers, extracted from social relations between partners, or based on roles and general properties of the agents). In addition, ReGreT has also a credibility module that is responsible of calculating how reliable the informants are. For the purpose of comparison we will only consider only Direct Trust and Witness Reputation (coming from informers).

5.5

The ReGreT calculation of Direct Trust is equivalent to the following predicate sequence in the Repage memory: Contract, Fulfilment, Outcome, Candidate Image and Image, with their corresponding detectors. In fact, the initial ideas for this part in the Repage memory were taken from ReGreT but adapted to the peculiar purpose of Repage. To calculate a new direct trust value, ReGreT uses a weighted aggregation of outcomes (where the notion of outcome is the same presented above, 4.13 for Repage).

5.6

The credibility module is used to calculate how reliable the informants are, and it is based on two types of information: the social relations among the agents involved and the accuracy of previous information coming from that informant.

5.7

Finally, the Witness reputation is calculated as a summation of the received information weighting each piece of information with the reliability of the informant as calculated by the credibility module.

Situation 1: The value of direct interaction

Repage

5.8

In this situation, *ag-0* receives a communication from *ag-I* saying that its image of *ag-T* as a seller is very good. At this moment *ag-0* has no image about *ag-I* as informer and therefore resorts to a default image that usually is quite low (see fig.4). The uncertain image as an informer adds uncertainty to the value of the communication. The result is that a quite defined evaluation (mostly very good, good) is transformed in a flat candidate image, that shows just a small prevalence of the good and very good values on the rest.
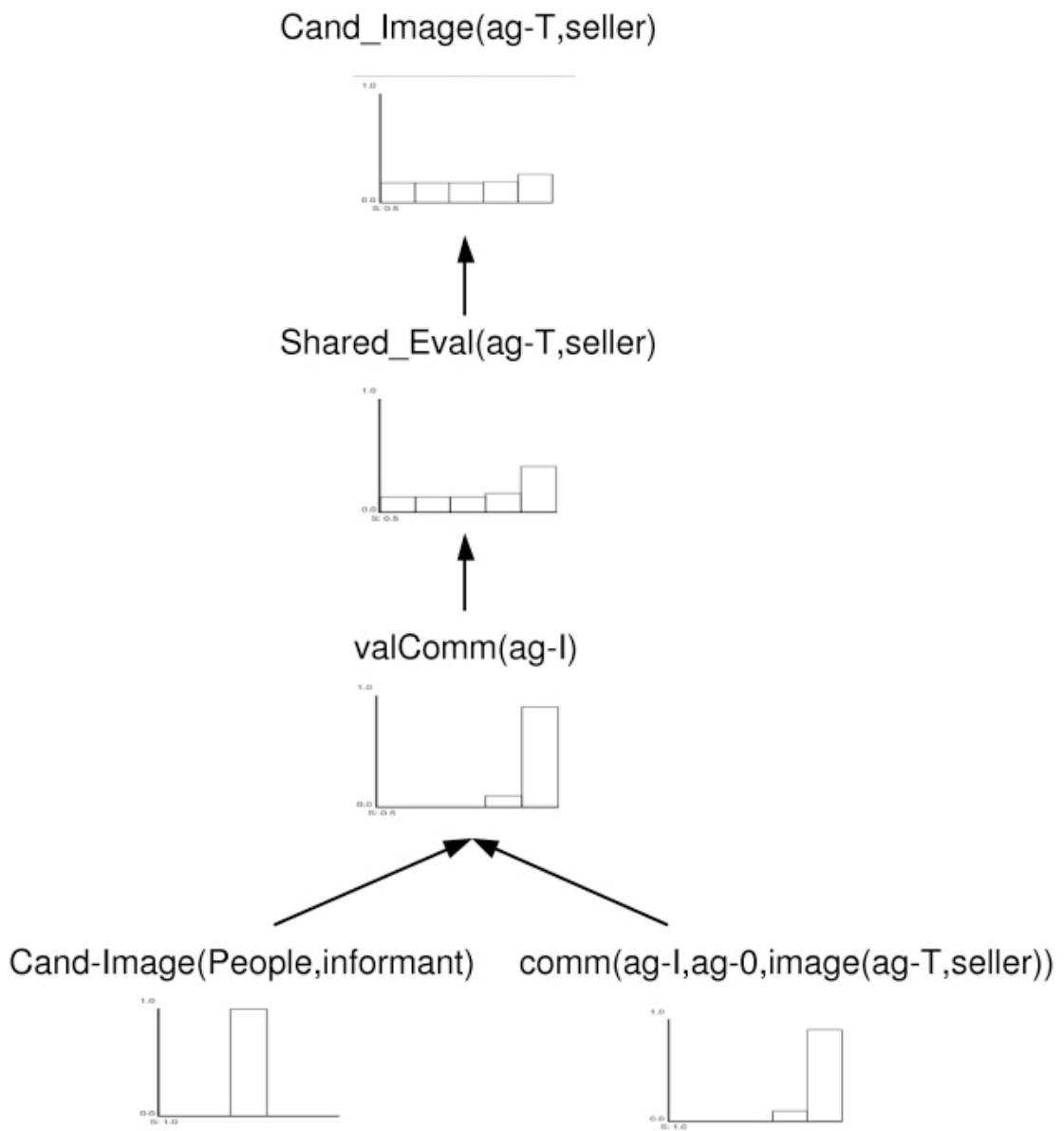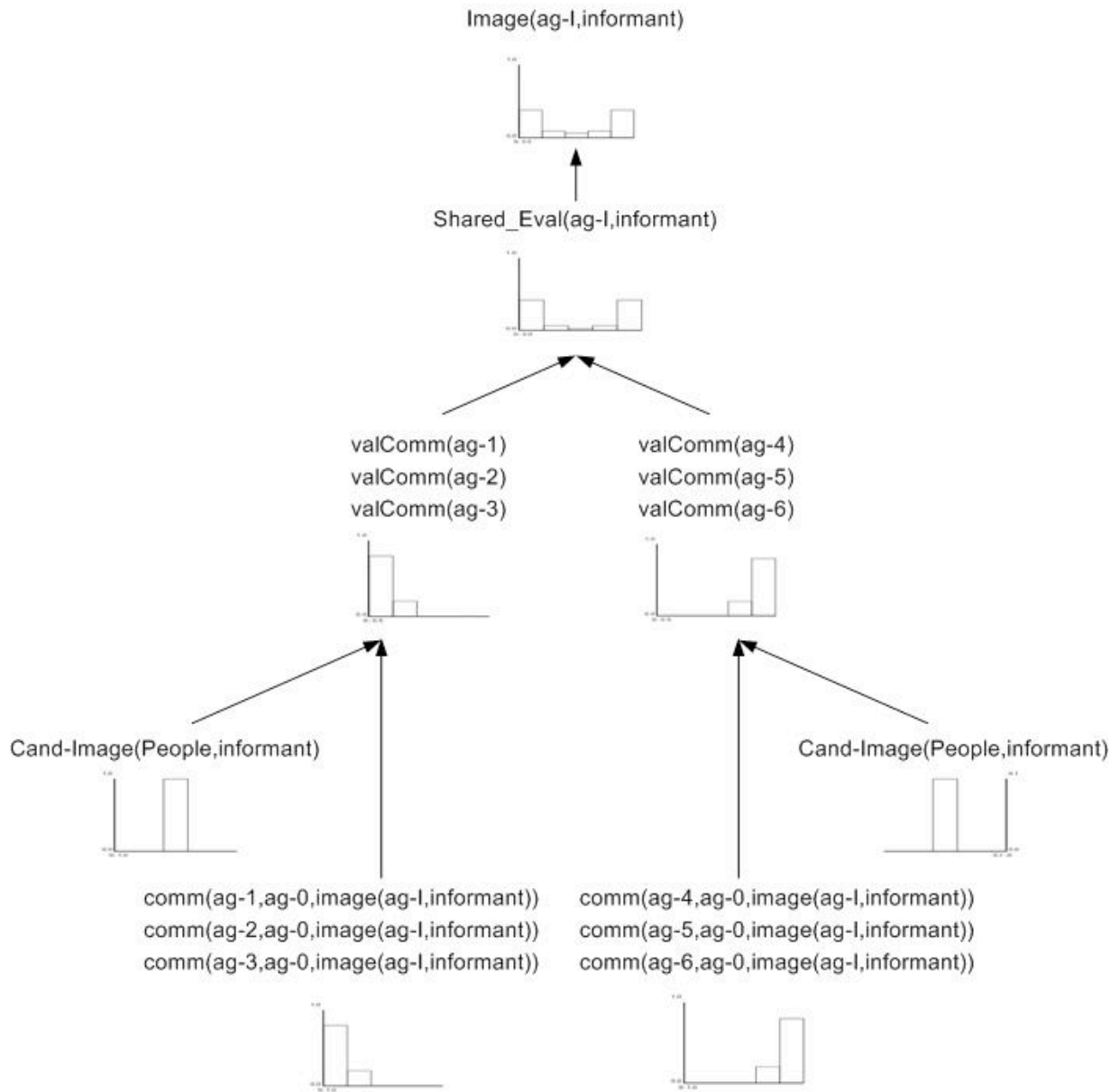
Cand_Image(ag-T,seller)

Shared_Eval(ag-T,seller)

valComm(ag-I)

Cand-Image(People,informant)   comm(ag-I,ag-0,image(ag-T,seller))

Figure 4. Value of direct interaction. Starting situation

5.9

Later on, *ag-0* has received 6 communications from different agents containing their image of *ag-I* as an informer. Three of them give a good report and three a bad one (see fig.5). This information is enough for *ag-0* to build an image about *ag-I* 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.

Image(ag-I,informant)

Shared_Eval(ag-I,informant)

valComm(ag-1)
valComm(ag-2)
valComm(ag-3)

valComm(ag-4)
valComm(ag-5)
valComm(ag-6)

Cand-Image(People,informant)

Cand-Image(People,informant)

comm(ag-1,ag-0,image(ag-I,informant))
comm(ag-2,ag-0,image(ag-I,informant))
comm(ag-3,ag-0,image(ag-I,informant))

comm(ag-4,ag-0,image(ag-I,informant))
comm(ag-5,ag-0,image(ag-I,informant))
comm(ag-6,ag-0,image(ag-I,informant))

Figure 5. Value of direct interaction. Arrival of contradictory information

5.10

At this point, *ag-0* decides to try a direct interaction with *ag-T*. Because he is not sure about *ag-T* he tries a low risk interaction. The result of this interaction is completely satisfactory and has important effects in the *Repage* memory. The candidate image about *ag-T* as a seller becomes a full image, in this case a positive one (see fig. 6).
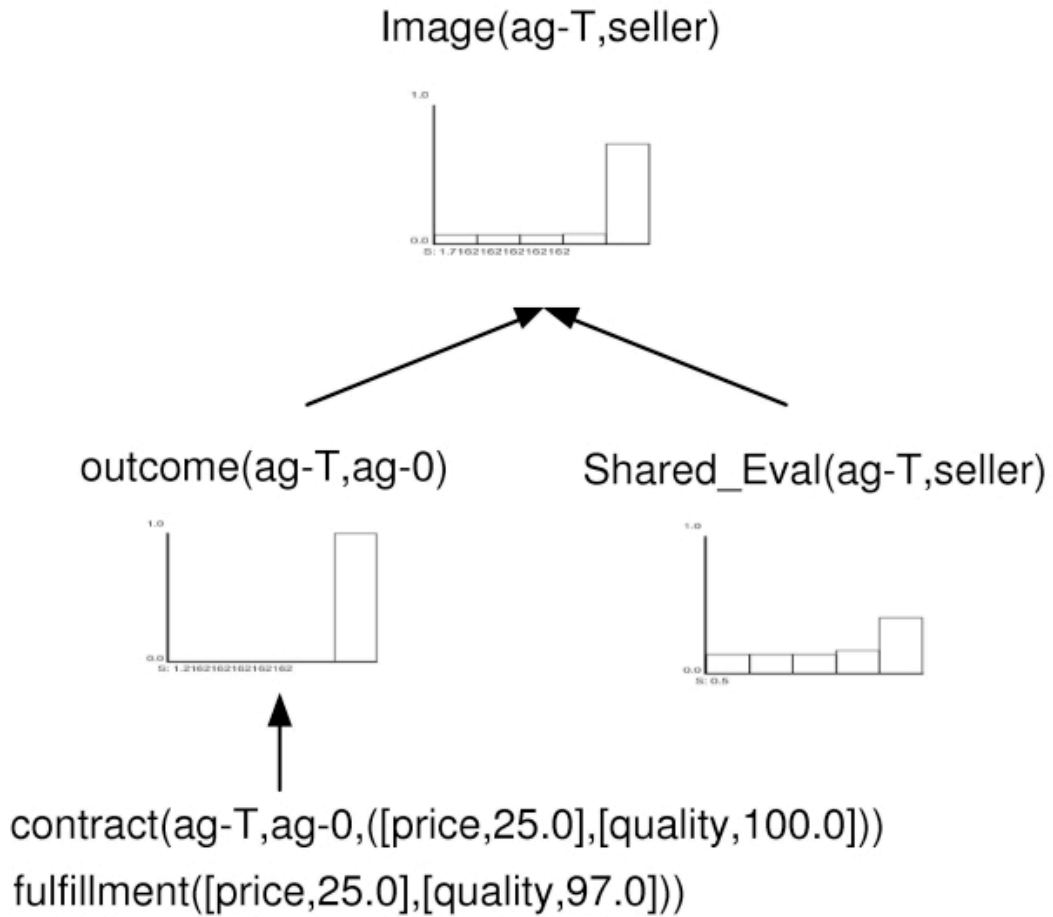
Figure 6. Value of direct interaction. Good direct interaction causes good image

5.11
Moreover, this positive image is compared (via the fuzzy metric) with the information provided by *ag-I* (that was a positive evaluation of *ag-T* as a seller); since the comparison shows that the evaluations are similar, a positive confirmation of the image of *ag-I* as an informer is generated. This reinforcement of the image of *ag-I* as a good informer at the same time reinforces the image of *ag-T* as a good seller (see fig. 7).
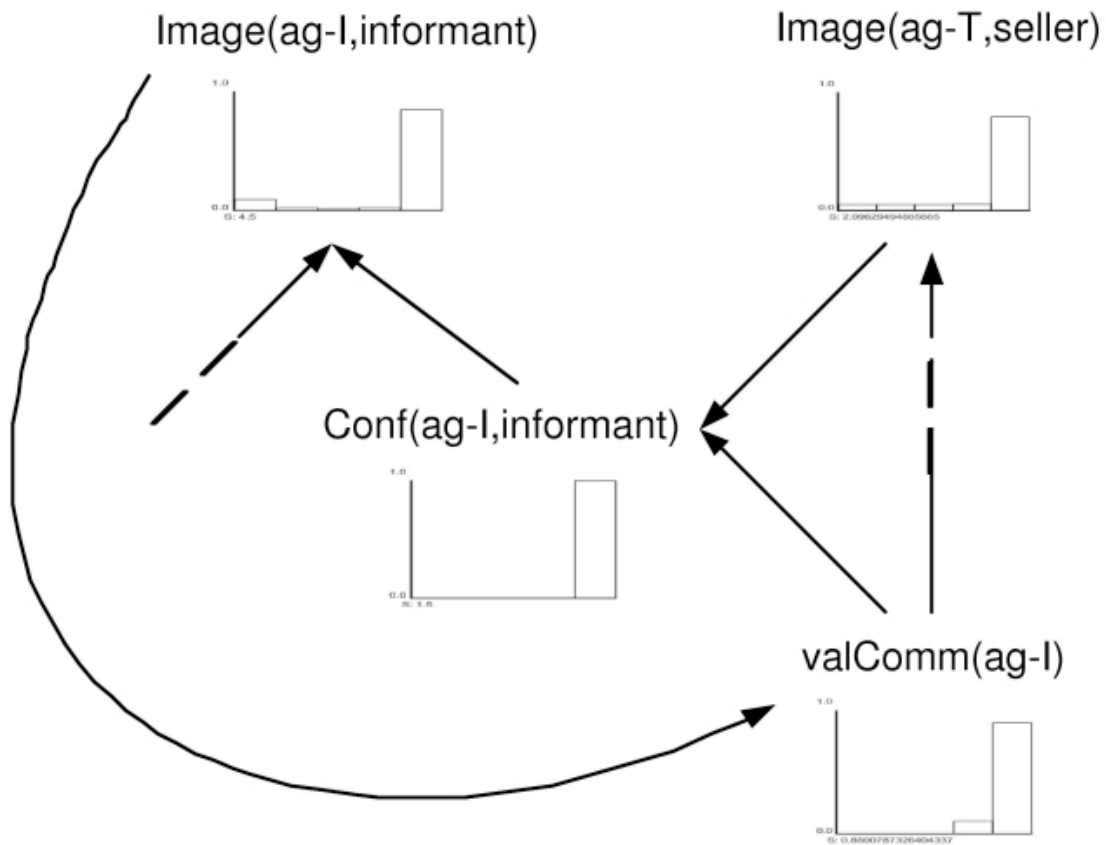
Figure 7. Value of direct interaction. Feedback loop

5.12

As a consequence, there is a positive feedback between the image of *ag-T* as a good seller and the image of *ag-I* as a good informer. This feedback is a necessary and relevant part of the model. However, only the first iteration is significant. The other iterations are weak "echoes" of the first one that have to be cancelled. Repage implements a threshold mechanism that allows the propagation of a predicate's value at time *t* only if there is a significant increment or decrement respect to the predicate's value at time *(t-1)*.

ReGreT

5.13

What would ReGreT do in the same situation? In this case, the results of the ReGreT model is quite similar to the one showed for Repage. Actually, ReGreT was one of the main influences in the Repage model and therefore they share several basic ideas.

5.14

The credibility model of ReGreT does not have any information about *ag-I* as informant. Therefore, it uses a default value to weight the relevance of the received information. This default value plays a similar role that the default candidate image in the case of Repage. With the arrival of the information about *ag-I* from different agents, ReGreT calculates a Witness Reputation of *ag-I* as informant. If we ask again ReGreT about the trust value of *ag-T* as a seller, the value will be calculated again taking into account the new information about *ag-I* as informant with a similar result of the Repage case. The final result will be also similar when the agent tries a direct interaction with *ag-T* and we make a new query to the ReGreT model about *ag-T* as a seller.

5.15

In this simple situation the final numerical results obtained if you use ReGreT or Repage (looking at them as black boxes) are very similar. However this similarity disappears when you look at the internals of the models. Repage has been designed to facilitate a meta–reasoning over the pieces of information that make possible the calculation of the final result; after the calculation, these predicates are kept in the memory and they will contribute to future elaboration with a much finer grain than the ones used in ReGreT or similar systems. Moreover, the architecture based on predicates that are linked among them allows not only to calculate the reputation values but also to analyze the origins of the results and reason about them.

Situation 2: Image is not reputation

Repage

5.16

The purpose of this situation is to show how Repage differentiates between image and reputation. In this case *ag-0*, after a couple of successful interactions with *ag-T* receives four communications from different informants. Each informant communicates the 'reputation' of *ag-T* as a seller. As we see in fig. 8, it happens that the reputation of *ag-T* is

negative but at the same time the image that *ag-0* has about *ag-T* is good (due to the direct interactions).
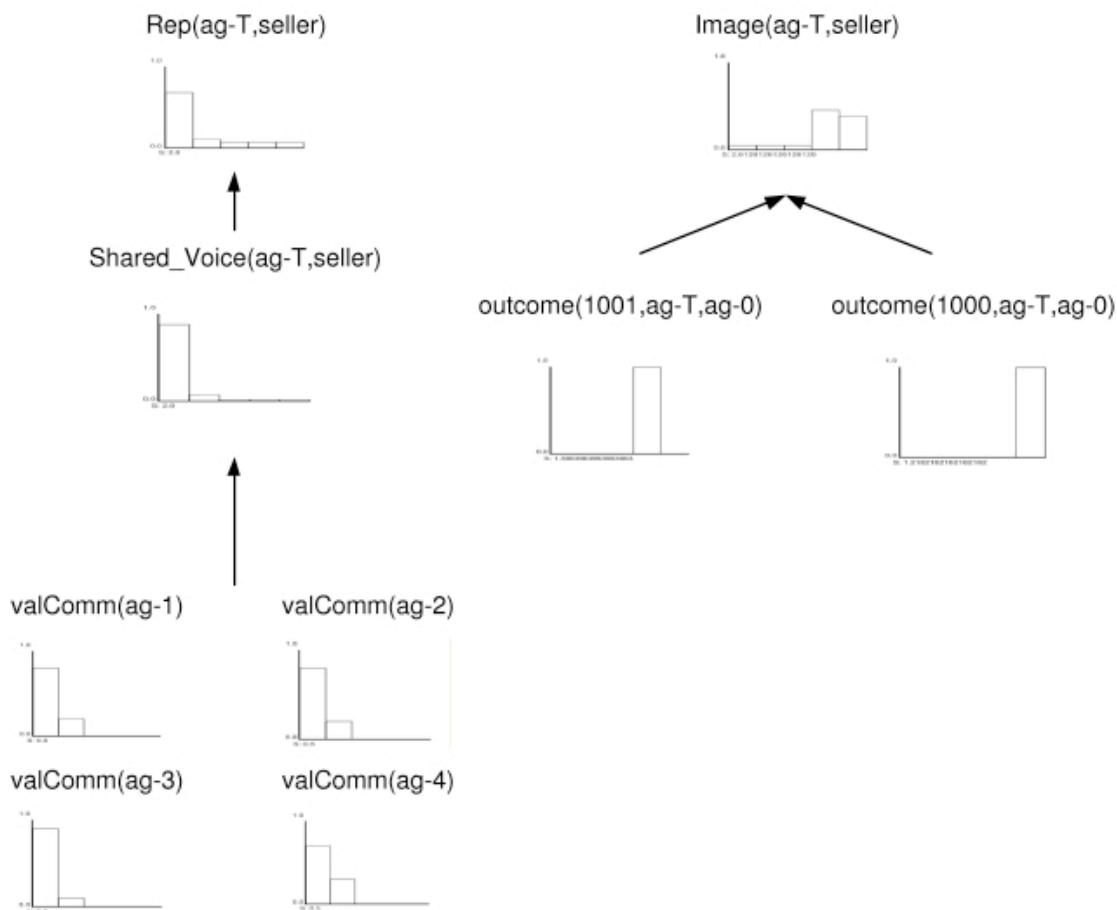
Figure 8. Image can be distinct from Reputation

5.17

That is not a problem from the point of view of Repage because there is a clear distinction between image and reputation. Also notice that, on the contrary of what happens with communicated images (see previous section), the communications about reputation as a seller do not generate confirmations that reinforce or weaken the image of the informant as an informant.

ReGreT

5.18

In this situation the differences between ReGreT and Repage are quite important. In ReGreT (and this can be generalized to other reputation models) there is no difference among image and reputation. ReGreT can differentiate initially the information concerning the reputation of ag–T as a seller (that contributes to the Witness reputation) from the outcomes or direct experiences (that contribute to the Direct Trust). However this distinction disappears once the agent starts to receive information concerning the image of *ag-T* as a seller from third party agents. ReGreT mixes both, reputation and image information. This means that an agent using ReGreT cannot keep track of the distinction between what it believes and what it believes the others believe.

Conclusions and future work

6.1

In this paper, the Repage system was presented as a tool for integrating image and reputation information in partner selection among autonomous but socially adaptable agents. The social cognitive theory on which the system is based was briefly described, and the system's architecture illustrated. The functioning of the system was exemplified by a set of situations showing how direct experience can prevail on contradictory communicated information, and the ability of the system to maintain separately reputation and image. As other systems developed in the MAS field, Repage is a modular and flexible tool, which helps selecting partners while taking into account others' experiences.

6.2

In addition, Repage has a number of advantages over similar systems. Based on a 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.

6.3

In the future, we plan several developments of the basic architecture. One promising direction is that of learning, in all of the memory manipulation modules. The analyzer at this moment is composed by a set of static rules. Also the biases

are static. Although this is enough for well-known and static environments, if we really want an "all-terrain" model we need some kind of learning and adaptation mechanism for these rules at run time. For example, although initially it seems reasonable to have a more positive bias toward outcomes than to witness information, if later witness information is revealed to be always accurate, this bias should be removed or at least reduced. The architecture of Repage has been designed having this necessity in mind.

6.4

Furthermore, artificial experiments comparing Repage with a mere image-based systems ought to be carried out in different domains of applications, including both cooperative settings like organisations and electronic marketplaces.

6.5

Finally, future developments of Repage ought to concern its integration with other components of a social agent, with special reference to learning and social adaptation, on one hand, and personalised inclinations and biases on the other.

References

BOLTON G E Katok E and Ockenfels A (2002) How effective are online reputation mechanisms? an experimental investigation. Max Planck Institute of Economics. Discussion Papers on Strategic Interaction, n. 25, https://papers.mpiew-jena.mpg.de/esi/discussionpapers/2002-25.pdf

CABRAL L M B and Hortaçsu A (2004) The Dynamics of Seller Reputation: Theory and Evidence from eBay, CEPR Discussion Papers 4345, http://econpapers.repec.org/paper/cprceprdp/4345.htm

CARBO J . Molina J M and Davila J (2002) Comparing predictions of SPORAS vs. a fuzzy reputation agent system. In *3rd International Conference on Fuzzy Sets and Fuzzy Systems, Interlaken*, pp. 147--153

CARBO J Molina J M and Davila J (2003) Trust management through fuzzy reputation. *Int. Journal in Cooperative Information Systems*, 12(1), pp.135-155

CONTE R and Paolucci M (2002) *Reputation in Artificial Societies: Social Beliefs for Social Order*. Boston: Kluwer.

CONTE R and Paolucci M (2003) Social cognitive factors of unfair ratings in reputation reporting systems. In *Proceedings of the IEEE/WIC International Conference on Web Intelligence — WI 2003*, pp. 316-322

DELLAROCAS C (2003) Efficiency and Robustness of Binary Feedback Mechanisms in Trading Environments with Moral Hazard, Paper 170, January 2003, Center for eBusiness@mit.edu

DUNBAR R (1998) *Grooming, Gossip, and the Evolution of Language*. Harvard Univ Press.

FALCONE R Pezzulo G and Castelfranchi C (2003) A fuzzy approach to a belief-based trust computation. *Lecture Notes on Artificial Intelligence*, 2631, pp. 73-86

GRUNIG J E and Hung C F (2002) The effect of relationships on reputation and reputation on relationships: A cognitive, behavioral study, Paper presented at the PRSA Educator's Academy 5th Annual International, Interdisciplinary Public Relations Research Conference.

HUYNH D Jennings N R and 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), New York, USA*, pp 65--74

KREPS D M and Wilson R (1982) Reputation and imperfect information. *Journal of Economic Theory*, 27, pp. 253-279

MICELI M and Castelfranchi C (2000) 'Human cognition and agent technology, in *The Role of Evaluation in Cognition and Social Interaction*. Amsterdam:Benjamins.

NOWAK M A and Sigmund K (1998) Evolution of indirect reciprocity by image scoring. *Nature*, 393, pp. 573-577

RAMCHURN S D Huynh D and Jennings N R (2004a) Trust in multiagent systems. *The Knowledge Engineering Review* 19 (1) pp. 1-25.

RAMCHURN S D Sierra C Godo L and Jennings N R (2004b) Devising a trust model for multi-agent interactions using confidence and reputation. *Int. J. of Applied Artificial Intelligence*, (18) pp. 833--852

RESNICK P and Zeckhauser R (2002) Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. In *The Economics of the Internet and E-Commerce*. Michael R. Baye, editor. Volume 11 of Advances in Applied Microeconomics. Amsterdam, Elsevier Science. http://www.si.umich.edu/~presnick/papers/ebayNBER/RZNBERBodegaBay.pdf

SABATER J and Sierra C (2002) Reputation and social network analysis in multi-agent systems. In *Proceedings AAMAS-02, Bologna, Italy*, pp. 475-482

SABATER J (2003) Trust and Reputation for agent societies. PhD thesis. Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Catalonia, Spain. http://www.iiia.csic.es/~jsabater/

SABATER J and Sierra C (2004) Review on computational trust and reputation models. *Artificial Intelligence Review*, vol. 24, no. 1, pp. 33-60

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

SKOWRONSKI J J and Carlston D E (1989) Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105, pp. 131–142

YAGER R (2004a) On the determination of strength of belief for decision support under uncertainity–Part I: generating strength of belief. *Fuzzy Sets and Systems*, 142, v. 1, pp. 117–128

YAGER R (2004b) On the determination of strength of belief for decision support under uncertainty–Part II: fusing strengths of belief. *Fuzzy Sets and Systems*, 142, v.1, pp. 129–142

YU B and Singh M P (2002) An evidential model of distributed reputation management. In *Proceedings of AAMAS-02, Bologna, Italy*, pp 294–301

ZADEH L A (1965) Fuzzy sets. *Inform. Control*, 8, pp. 338–353