Web Intelligence 13 (2015) 295–312
DOI 10.3233/WEB-150327
IOS Press

On the emergence of semantic agreement among rational agents

Golnaz Vakili [a,*], Thanasis G. Papaioannou [b] and Karl Aberer [b]
[a] *Iranian Research Institute for Information Science and Technology, Iran*
E-mail: vakili@irandoc.ac.ir
[b] *Ecole Polytechnique Fédérale de Lausanne, Switzerland*
E-mails: thanasis.papaioannou@epfl.ch, karl.aberer@epfl.ch

Abstract. Today's complex online applications often require the interaction of multiple (web) services that belong to potentially different business entities. Interoperability is a core element of such an environment, yet not a straightforward one due to the lack of common data semantics. The problem is often approached by means of standardization procedures in a top-down manner with limited adoption in practice. (De facto) standards for semantic interoperability most commonly emerge in a bottom-up approach, i.e., involving the interaction and information exchange among self-interested industrial agents. In this paper, we argue that the emergence of semantic interoperability can be seen as an economic process among rational agents and, although interoperability can be mutually beneficial for the involved parties, it may also be costly and might fail to emerge. As a sample scenario, we consider the emergence of semantic interoperability among rational web service agents in service-oriented architectures (SOAs), and we analyze their individual economic incentives with respect to utility, risk and cost. We model this process as a positive-sum game and study its equilibrium and evolutionary dynamics. According to our analysis, which is also experimentally verified, certain conditions on the communication cost, the cost of technological adaptation, the expected mutual benefit from interoperability, as well as the expected loss from isolation, drive the process.

Keywords: Semantic interoperability, incentives, game theory, schema mappings, consensus

1. Introduction

Today's emerging complex online applications often require the orchestration of multiple web services that potentially belong to different business entities and can be assumed to be provided by autonomous agents. According to the service-oriented architecture (SOA) design principles, services should be flexible in order to discover, select, and use other services for fulfilling a given task and goal [24]. Service interaction can be achieved by syntactic interoperability, whereas conformance to the service roles in the service orchestration pertains to semantic interoperability [5]. Semantic interoperability refers to the ability to exchange data with unambiguous shared meaning. Standardization efforts have a long history dealing with syntactic interoperability, whereas semantic interoperability is still open and mainly dealt with by ontology alignment, schema mapping and attribute correspondence approaches [40]. Apart from the technical challenges for achieving semantic interoperability, it often emerges as a bottom-up dynamic decision-making process among self-interested industrial agents, and it involves costly information exchange and technological adaptation for conformance to the agreed standards [1]. Bottom-up approaches have been already proposed in the web-service community, by studying the notion of folksonomies, or emergent behavior of agents. In this era, it is often assumed that standardization means conformance to W3C or other similar formats, such as XML Schema, WSDL, SOAP, BEPL4WS, rather than the notion of ontological agreement or canonical dictionaries for given domains. The benefits from adopting standards [10], as well as the

*Corresponding author. E-mail: vakili@irandoc.ac.ir.

decision-making process regarding the optimal timing for the adoption of a new standard [12], have also been well studied. However, to the best of our knowledge, the dynamic process of reaching a distributed agreement towards a standard among rational agents has not been analyzed.

In this paper, we employ a game-theoretic approach to analyze the individual incentives among rational agents for decision-making towards the emergence of semantic agreement.[1] As a running example, we consider the real problem of achieving semantic interoperability for web-service composition in the SOA context among rational service agents that employ different schemas (or ontologies) to describe their interfaces. Service agents that employ a common schema (or ontology) and seek to utilize services from other agents that employ a different schema (or ontology) have to collectively reach local semantic mappings (or alignments) between the two schemas (or ontologies) of acceptably-high quality, through costly iterative communication and mapping re-adjustment. If a semantic agreement is reached (i.e., semantic mappings of acceptably-high quality have been found), a benefit is expected to be mutually enjoyed by the service agents involved in the semantic mapping creation process, whereas a loss is expected for the rest of the agents due to their semantic isolation. We model this process as a constant positive-sum game and study its evolutionary dynamics in order to find the conditions upon which semantic agreement can emerge as a bottom-up process among rational agents. We identify as the most important parameters for the emergence of an agreement on semantic mappings: (i) the expected benefit from achieving semantic interoperability, (ii) the cost of technological adaptation to the agreed semantic mapping, (iii) the expected semantic isolation cost, and (iv) the minimum required quality of the schema mapping for achieving agreement. The conditions for reaching semantic agreement as an evolutionary stable strategy (ESS) are identified and verified by simulation experiments. Our modeling can be applied in more general distributed settings for achieving semantic interoperability among rational agents.

The remainder of this paper is organized as follows: In Section 2, we describe our running example. In Section 3, we model the decentralized decision making on adopting/rejecting semantic mappings as a game and then find the conditions for the beneficial equilibrium. Also, we model the decentralized decision-making process for advertisement of partial semantic mappings. In Section 4, we revisit the game model of Section 3 and analyze it as a Bayesian game according to the updated belief on reaching consensus. In Section 4.3, we discuss on equilibrium selection in the Bayesian game for semantic convergence. In Section 5, we numerically evaluate our results (Section 5.1) and we verify our analytical results with simulation experiments (Section 5.2). In Section 6 we discuss the related work and finally, in Section 7, we conclude our work.

[1] A short and preliminary version of this work was published in [39]. The current work has been significantly enriched with a comprehensive running example, with improved analysis and result proofs, with comprehensive experiments and with a more detailed review of the related work.

2. The running example

We focus on a specific problem in semantic interoperability as a sample scenario that will be used as a running example throughout the paper. The purpose of this sample scenario is to address similar semantic interoperability problems encountered in real settings and applications of web services (e.g., [23,41]) or among different business entities. We consider a system of distributed service repositories where service providers register their service interfaces. At a registry, service descriptions are represented, based on a specific structured schema. Our problem is how to achieve semantic interoperability in a web service composition example.

Interoperability between any two sets of service providers registered in two different registries requires the existence of *semantic mappings* (or *alignments*) between their corresponding schemas (or ontologies), e.g., XML schemas. For example, to constitute a value-added social-event ticketing service, a profile management web service at the first registry is combined with a ticket management service and a pdf generation service at the second registry, as depicted in Fig. 1(a). Thus, semantic interoperability increases the business potential for the service agents at either registry. A sample mapping (alignment) between the ontologies of Registries 1 and 2 is depicted in Fig. 1(b). The formalism of such schema mappings is beyond the scope of our work and, without loss of generality, simple attribute correspondences can be assumed [16]. For example, (Purchase.person.User.firstname = Transaction.customer.Customer.name), or (Purchase.when = Transaction.date) that are given as XML path cor-

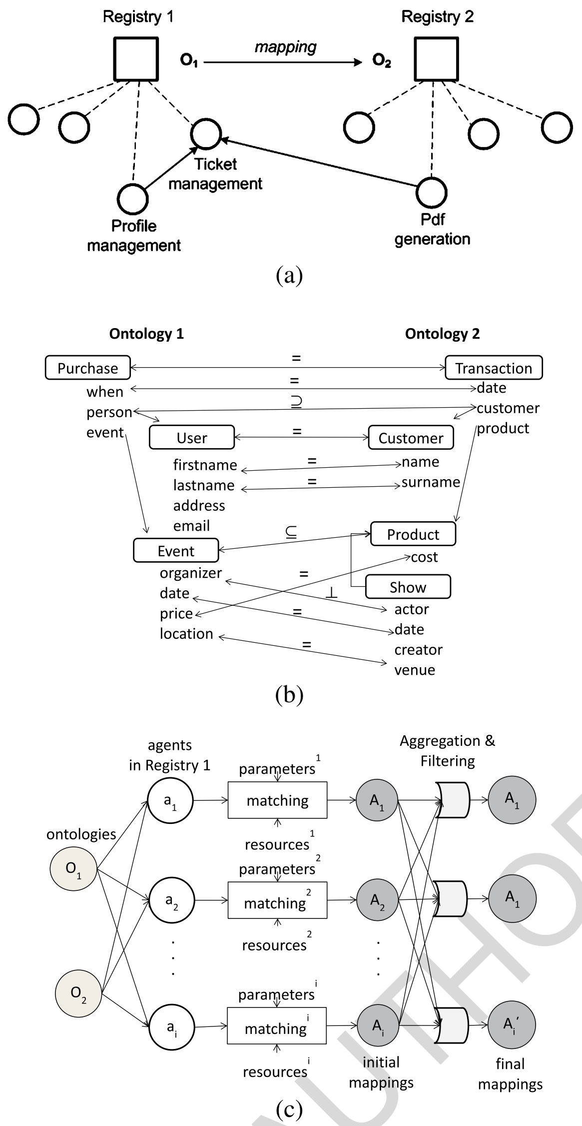

Fig. 1. a) The sample scenario. b) A sample mapping. c) The matching process.

respondences (the dot syntax can be assumed either an is-a relationship or a subclass relationship) are instances of attribute correspondences between the ontologies in Fig. 1(b). More formally, in the ontology alignment domain [16], a correspondence (or partial mapping here) between two ontologies O and O', is a 5-uple $< id, e, e', r, n >$ such that id is an identifier of the correspondence, e and e' are entities of O and O' (e.g. XML elements, concepts), r is a relation (e.g. equivalence ($=$), more general ($>=$), disjointness ($! =$)), and n is a confidence measure typically in $[0, 1]$. Semantic mapping or alignment between ontologies O and O' is a set of correspondences (or partial mappings) on O and O', with some additional metadata (e.g. multiplicity: 1–1, 1–*, method, date, properties, etc.). By a semantic mapping between the schema of the first registry to that of the second one, the service agents at the first registry can employ the services of the agents at the second registry.

Any subset of the service providers that are registered at the same registry can communicate to each other, for example in ebXML architecture [14]. Semantic interoperability in this setting is established by a consensus among the set of service providers, registered at one registry, on the acceptable quality of the schema mapping that is going to be used in their collaboration with another set in the system. This consensus is mutually beneficial for all the service agents at the registry, as the more the agents that reach a schema mapping of high quality are, the higher the business potential for collectively constructing complex value-added services with service agents from the second registry is.

As a matching process, we assume a parallel composition scenario (depicted in Fig. 1(c)), as follows: Initially, each agent employs a separate matcher (e.g. S-Match, OLA, COMA++, etc.) with its own parameters (provided by experts) and resources, and it produces an individual mapping of limited effectiveness. Then, agents exchange partial mappings with each other (through costly communication) and gradually converge to a new mapping of improved quality through aggregation and filtering of partial mappings. In the aggregation and filtering process, we assume that each agent, upon receiving a partial mapping, is able (using one of the quality metrics in [17]) to evaluate whether replacing its corresponding local partial mapping leads to an improvement in the schema mapping quality. Moreover, a certain partial mapping might violate a 1-to-1 constraint in the local schema mapping or lead to an inconsistency by creating incorrect attribute-correspondence circles. On one hand, adopting a new partial mapping might involve future costs for the agent (e.g. for the implementation and the system integration of this mapping), when the schema mapping is materialized. On the other hand, keeping a heterogeneous schema mapping can result in market losses due to semantic isolation. In this economic process, two different equilibria may emerge depending on the agents' decisions. In the first one, the agents keep their heterogeneous schema mappings of low quality, as they fail to collaboratively build schema mappings of high quality, as opposed to the second one, where they achieve semantic interoperability by having reached a consensus on schema mappings of acceptably-high quality. Thus, the emergence of semantic interoperability can be viewed as a decentralized coordination problem for finding an *agreement* among rational agents on the quality of their mapping to another schema.[2] The complexity of the coordination problem results from imposing two important constraints: (i) The service agents are self-interested and profit seekers; and (ii) they exchange partial mappings through local communications (i.e., limited-scope advertisements of partial mappings). To tackle the problem, we develop a decentralized strategy in order to coordinate autonomous agents, with respect to their rationality and local communication. Then, we investigate under which conditions the consensus does emerge in such an economic process.

3. Decentralized strategy

In this section, we develop a decentralized strategy to coordinate the rational decisions of autonomous agents for collectively constructing a schema mapping of acceptable quality based on local communications. We denote as P the service agents that are registered at a registry R and they use the schema S to express their services. An advertised partial mapping μ can be understood as an attribute mapping between the source schema S and a destination schema S'; e.g. $A \rightarrow B$ and $A \rightarrow C$ are two instances of partial mapping μ, where $A \in S$ and $B, C \in S'$.

We refer to the fraction of the agents that reach agreement on a certain minimum acceptable quality of a schema mapping as the *consensus level* for that schema mapping. The consensus level is generally unknown, and thus we model it as a random variable $\Phi \in [0, 1]$. As different instances of the schema mapping have different qualities, we also model the quality of the schema mapping with another random variable $Q \in [0, 1]$. The probability distribution of these random variables is discussed later in this section. To facilitate the decision making analysis, we employ a discrete random variable Θ on the consensus state that is

Table 1

The notation used in developing the decentralized strategy

Symbol	Description
Q	The random variable associated with the quality of the schema mapping
Θ	The random variable on the *consensus* state
Φ	The random variable associated with the *consensus level*
ϕ_0	The *consensus level threshold* that is sufficient for the *agreement*
a_1, a_2	Two choices upon receiving an instance of mapping: a_1: *reject*, a_2: *adopt*
b	The marginal benefit of adopting the received instances of; mapping if a *consensus* is finally established
c_a	The adoption cost incurred for schema mapping modification, implementation and system integration
c_h	The heterogeneity cost due to the market losses for an agent that does not possess a high quality mapping as compared to other agents
γ	The ratio of the adaptation cost to the sum of the benefit and the heterogeneity cost
c_{adv}	The cost incurred for each time of advertising an instance of the mapping
X	The random variable associated with the estimated current quality of the schema mapping of other agents
$r(.)$	The risk associated with the advertisement or not of a particular mapping

based on Φ, as follows:

$$\theta_\phi = \begin{cases} \theta_1; & \phi \in [0, \phi_0) \\ \theta_2; & \phi \in [\phi_0, 1] \end{cases} \tag{1}$$

where ϕ_0 is the *consensus level threshold* that is sufficient for the *agreement*. According to the definition of θ, the set of all possible states of the uncertain parameter is θ_1, *no consensus* and θ_2, *consensus*. Therefore, we define that a consensus on quality of the schema mapping is established when at least ϕ_0 proportion of the service agents P reach an agreement on a certain minimum quality of it. The decentralized strategy will be developed in the following subsections: to guide the decisions of autonomous agents on advertising their partial mappings, on one hand, and on adopting the received advertisements, on the other.

3.1. Adoption

The received instance of a mapping μ is taken into consideration by a rational agent if it results in a schema mapping of higher quality, as compared to the one currently used (by employing one of the quality metrics in [17], e.g., average string similarity); other-

[2]Note that, after agreement, agents might still possess different mappings, but of individually acceptable qualities. Different agents might have different preferences regarding the acceptable quality of their schema mapping that they rationally pursue.

wise it will be ignored without further decision analysis. To guide the rational decisions of the agents, the decentralized strategy is developed based on a decision model that is constructed in accordance with the expected individual utility maximization. Each agent has two choices upon receiving an instance of mapping μ that results in a more qualified schema mapping: a_1: *reject*, or a_2: *adopt* it. The uncertain parameter that affects the decision process is the state of consensus on quality of the schema mapping. Then, an agent will gain a *utility* $u(a|\theta)$ according to its action a and depending on the eventually-realized state of the consensus θ.

Semantic interoperability is considered to be mutually beneficial for the service agents if they agree on an acceptable quality of the schema mapping; thus, b is defined to represent the marginal benefit of adopting each of the received instances of μ if a consensus is finally established.[3] Meanwhile, a costly effort c_a should be exerted by the agent for the eventual implementation and system integration of the partial mapping μ. In general, c_a can be different for different partial mappings, but we assume it to be constant for modeling simplicity. Also, in general, b and c_a can be different for different agents for the same mapping, because, for example, high-quality schema mappings might have different benefits for different agents, and more or less effort might be required for the implementation of a particular partial mapping. We experimentally consider this case in Section 5. If an agent adopts the received instance of partial mapping μ and a consensus is reached, he gains a utility $b - c_a$; otherwise, the outcome is only a loss of the value c_a, as he still incurs the adoption cost for attempting interoperability (i.e., schema mapping modification, schema mapping implementation design), yet without getting any benefit from it. On the contrary, if an agent rejects a mapping that others have collectively found it acceptable, then while he should not expect any benefit from interoperability, he should also pay a *heterogeneity cost*; this cost results from his expected market losses from his semantic isolation. We denote c_h as the marginal cost of heterogeneity expected to be incurred. Last, in the case that an agent rejects a received instance of μ and there is eventually no consensus, then no cost or benefit occurs and

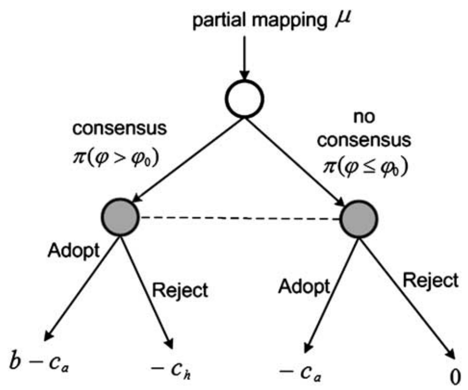

Fig. 2. The strategic form of the game.

thus no utility is obtained. To summarize, the utility function[4] u from the adoption decision process is given by

$$u(a|\theta) = \begin{cases} b - c_a; & a = a_2, \theta = \theta_2 \\ 0; & a = a_1, \theta = \theta_1 \\ -c_h; & a = a_1, \theta = \theta_2 \\ -c_a; & a = a_2, \theta = \theta_1 \end{cases} \qquad (2)$$

This decision-making process can be modeled as a positive-sum game among rational agents with respect to their derived utilities. For clarity, the strategic form of this game is depicted in Fig. 2.

Based on Nash equilibrium analysis, there will be two equilibria[5] constituted in $\{(a_2^i, a_2^{-i})|\theta_2\}$ and $\{(a_1^i, a_1^{-i})|\theta_1\}$ where a^i is the action of an individual p_i and a^{-i} is the action of the other agents $p_j \in P$, $p_j \neq p_i$. Although the first equilibrium is Pareto-efficient and achieves semantic interoperability (as in this case, a certain (ϕ_0) proportion of agents reach an agreement on an acceptable quality of the schema mapping), a natural question arises about which equilibrium will be chosen by the system. It depends on the state of θ and the probability distribution of $\pi(\theta)$ that should actually be obtained. It can be concluded from maximum entropy priority[6] [9] that

[3]The marginal benefit b can be calculated by dividing the total benefit of having a high-quality mapping divided by the estimated number of mappings that should be adopted for reaching the acceptable quality.

[4]The utility function of agents is the base for their "iterative" decision making process, and hence it does not capture one-time costs paid by all the agents, such as the cost of generating the mapping. This kind of cost is part of the initialization costs and does not affect the dynamics of the process.

[5]A Nash equilibrium is an action profile with the property that no single player can obtain a higher payoff by deviating unilaterally from this profile.

[6]In Bayesian probability, the principle of maximum entropy is a postulate that states that, subject to known constraints (called

$\pi(\phi) = U[\alpha, \beta]$ where $\alpha = 0$ and $\beta = 1$, in order to model a prior distribution for ϕ as non-informative as possible. Then, the probability distribution of θ is given by

$$\pi(\theta) = \begin{cases} \phi_0; & \theta = \theta_1 \\ 1 - \phi_0; & \theta = \theta_2 \end{cases}$$

According to the Bayes decision principle [9], a *rational* agent's decision maximizes $\rho(a, \pi)$, where ρ is the Bayesian expected utility and is defined as

$$\rho(a, \pi) = E^{\pi}\big[u(a|\theta)\big] = \sum_{\theta_i \in \Theta} u(a|\theta_i)\pi(\theta_i) \quad (3)$$

Therefore, to make an optimal decision and choose the action a that maximizes $\rho(a, \pi)$, the expected utility of a_1: *reject* and a_2: *adopt* is assessed respectively by each agent:

$$\begin{aligned} E^{\pi}\big[u(a_1|\theta)\big] &= u(a_1|\theta_1)\cdot \pi(\theta_1) + u(a_1|\theta_2)\cdot \pi(\theta_2) \\ &= 0 \times \phi_0 + (-c_h) \times (1 - \phi_0) \\ &= -c_h(1 - \phi_0) \\ E^{\pi}\big[u(a_2|\theta)\big] &= u(a_2|\theta_1)\cdot \pi(\theta_1) + u(a_2|\theta_2)\cdot \pi(\theta_2) \\ &= (-c_a) \times \phi_0 + (b - c_a) \times (1 - \phi_0) \\ &= b(1 - \phi_0) - c_a \end{aligned}$$

Finally, setting $E^{\pi}[u(a_1|\theta)] = E^{\pi}[u(a_2|\theta)]$ will result in

$$\frac{b + c_h}{c_a} = \frac{1}{1 - \phi_0} \quad (4)$$

For convenience, we define gamma as the ratio of the adaptation cost of a partial mapping to the sum of its benefit and the heterogeneity cost, i.e.,

$$\gamma \equiv \frac{c_a}{b + c_h}$$

To this end, it can be concluded that, when a received instance of mapping μ is taken under consideration by a rational agent $p_i \in P$, it is *rejected* if $\gamma_i > (1 - \phi_0)$ and *adopted* if $\gamma_i < (1 - \phi_0)$; where threshold of ϕ_0 can be defined to have the same value for all the participating agents or it can be set by individual agents to different values.[7] If the decision analysis results in a_1 as the Bayes action in more than $(1 - \phi_0)$ percent of agents, then it is impossible to reach an agreement on an acceptable quality of the schema mapping; because in this case, with reference to (1), the proportion of the agents that would prefer to keep and use their current instances of the schema mapping would be too high to achieve a consensus on the quality. Therefore:

Result I. *The necessary condition for existence of the Pareto-efficient equilibrium in the system is that at least ϕ_0 percent of agents $p_i \in P$ conclude $\gamma_i < (1 - \phi_0)$.*

The interpretation of this result is that semantic interoperability does emerge in the system only if, in ϕ_0 percent of the agents $p_i \in P$, the γ_i (the ratio of adoption cost to the cost and benefit of interoperability) is less than $1 - \phi_0$. To clarify this through an example, suppose a collaboration scenario in which, the threshold ϕ_0 is determined to be 0.75 according to the requirements of the services provided by the agents, then for at least 75 percent of the rational agents, the incentive of interoperability (in terms of their b and c_h) should be four times their adoption cost; only then it would be possible to achieve a consensus.

3.2. Communication

In this section, we develop the advertisement decision-making model of the decentralized strategy based on individual risk minimization. As will be discussed, the advertisements of partial mappings significantly affect the agents' local perception of the level that the global consensus has been reached, hence their individual Bayesian decision process for the adoption of the partial mappings that they receive.

There is always a risk (defined as the expected loss in [0, 1]) associated with the advertisement of a partial mapping. The advertised partial mapping will be ignored by the receiver agents, if it does not result in a more qualified schema mapping. In such a case, the agent has a cost of c_{adv} for the required communication and for externalizing internal knowledge.[8] A receiving cost is also anticipated in the communication cost, however here, we focus on the sending cost, because this is the dominant cost component. There is also an

testable information), the probability distribution that best represents the current state of knowledge is the one with largest entropy.

[7] We experimentally investigate the latter setting in the evaluation section.

[8] Such a cost would be reasonable (under rationality assumptions) for locally-generated mappings that incur a certain generation cost.

opportunity cost from not advertising a given mapping; this cost is due to the probability of not reaching a consensus due to not advertising a partial mapping of high quality.[9] In other words, if a consensus is expected to be reached, then the advertisement cost is expected to be a justified investment cost. In the case that a consensus is not expected to be reached, the adoption cost c_a for the current instance of the local partial mapping is considered to be a loss. Thus, the risk of advertising can be modeled as the probability of not reaching the consensus, and the risk of not advertising can be modeled as the probability of reaching the consensus. Assuming that the agents are risk averse with respect to their advertisement actions, if the expected loss from not advertising a certain partial mapping is greater than that due to the advertisement, then the agent should advertise it. Formally, assuming a decision space $\Delta = \{\delta_2, \delta_1\}$ for advertisement, where δ_2 and δ_1 respectively correspond to advertising or not a given partial mapping μ, the agent chooses δ_2 if the following condition is true:

$$c_{adv} \cdot r(\delta_2) < c_a \cdot r(\delta_1) \tag{5}$$

where c_{adv} is the advertisement cost, c_a is the adoption cost, and $r(.)$ is the risk associated with the advertisement or not of a particular partial mapping.

According to Eq. (5), the agents decide to advertise a partial mapping or not based on their local belief on the probability of reaching the consensus. In this section, we assume this belief to be fixed, regardless of the information that becomes available to the agent on the achieved level of the consensus. If π denotes the probability of reaching the consensus, then it is $r(\delta_1) = \pi = 1 - r(\delta_2)$; thus, the advertisement policy that is derived from (5) is fixed over time. More specifically, as we found in the previous section, the probability of reaching the consensus is $\pi = 1 - \phi_0$. Evaluating $r(\delta_1)$ and $r(\delta_2)$ in Eq. (5), the optimal advertisement rule δ^* is determined with respect to ϕ_0 by the following formula:

$$\delta^* = \begin{cases} \delta_1; & \frac{c_{adv}}{c_a} > \frac{1-\phi_0}{\phi_0} \\ \delta_2; & \frac{c_{adv}}{c_a} < \frac{1-\phi_0}{\phi_0} \end{cases} \tag{6}$$

Note that, if $c_{adv}/c_a = 1 - \phi_0/\phi_0$, then the agent is indifferent between advertising or not and randomly selects δ_1 or δ_2. To this end, advertising or not the partial schema mappings depends on the relative costs c_{adv}, c_a and the consensus level threshold ϕ_0.

[9] In the rest of paper, by quality of a partial mapping we mean, the quality of the schema mapping that results from making use of this partial mapping.

4. Bayesian analysis of decentralized strategy

In this section, we will demonstrate how the developed strategy provides the sufficient condition and reduces the already found necessary condition for the emergence of the Pareto-efficient equilibrium. This is achieved by employing the information that becomes locally available through advertisement of partial mappings in the individual decisions of the autonomous agents.

4.1. Adoption

In this section, we employ the conditional Bayes decision principle in the decision model of the decentralized strategy. According to this principle, every agent assesses its expected utility from the adoption or the rejection of an advertisement, with respect to the believed distribution of ϕ at the time of decision making. Let us consider the snapshot of the estimated current quality of the schema mapping of the rest of agent community based on the received sample of the partial mappings. This quality snapshot can be seen as a random variable X with probability distribution $f(x)$ that depends on the consensus level ϕ.

Theorem I. *The* quality *Q and the* consensus level Φ *of the schema mapping are* equal in distribution.

Proof. Agents advertise mappings of high quality among the ones that they possess. Recall that the consensus level Φ is the fraction of the agents that have reached a minimum quality threshold. Therefore, a low average quality x of advertised mappings reveals a low fraction of agents that have reached a minimum acceptable quality. As x increases, so does the consensus level, and hence Q and Φ are equal in distribution:

$$\forall x \in [0, 1], \quad P(Q \leqslant x) = P(\Phi \leqslant x) \tag{7}$$

□

Taking advantage of the information that becomes available through advertisements, the distribution $\pi(\phi)$ becomes the posterior distribution $\pi(\phi|x)$. This distribution reflects the updated belief of a participant on the global consensus level of the schema mapping, after observing the sample x through the advertisement of partial mappings.

To this end, the uniform distribution of ϕ over 0 and 1 is revised to model $\pi(\phi|x)$. The upper bound of the posterior distribution remains 1. However, the lower bound of $\pi(\phi|x)$ is set to α, where based on the equality of Φ and Q in distribution, α can be considered by individual agents as the minimum quality of the schema mapping after employing the received advertisements. Therefore, the believed probability distribution of θ at the time of decision making is assessed by individual agents as follows:

$$\pi^*(\theta) = \begin{cases} (\phi_0 - \alpha)/(1-\alpha); & \theta = \theta_1 \\ (1-\phi_0)/(1-\alpha); & \theta = \theta_2 \end{cases}$$

where $\pi(\theta|x)$ is denoted by $\pi^*(\theta)$. Based on the value of α, two different cases can be distinguished, specifically:

(i) if $0 < \alpha < \phi_0$, then the expected utilities are assessed in accordance with Bayesian expected utility formula:

$$\begin{aligned} E^{\pi^*}\big[u(a_1|\theta)\big] &= -c_h(1-\phi_0)/(1-\alpha) \\ E^{\pi^*}\big[u(a_2|\theta)\big] &= b(1-\phi_0)/(1-\alpha) - c_a \end{aligned} \tag{8}$$

With reference to Bayes decision principle (in Section 3.1), an optimal decision maximizes $\rho(a, \pi^*)$. $E^{\pi^*}[u(a_1|\theta)]$ and $E^{\pi^*}[u(a_2|\theta)]$ are *monotonic* functions of α that are *strictly* decreasing and increasing respectively. Therefore, to determine the Bayes action a^{π^*}, we should find the crosspoint at which $E^{\pi^*}[U(a_1|\theta)] = E^{\pi^*}[U(a_2|\theta)]$. From the latter, after some algebra, we obtain that

$$(1-\alpha)\gamma = (1-\phi_0) \tag{9}$$

Thus, in this case, when a received instance of a partial mapping is considered by a rational agent $p_i \in P$, it is *rejected* if $\gamma_i\,(1-\alpha_i) > (1-\phi_0)$ and *adopted* if $\gamma_i\,(1-\alpha_i) < (1-\phi_0)$. As discussed in Section 3, if the decision analysis results in $a^{\pi^*} = a_1$ in more than $(1-\phi_0)$ percent of service agents, it is impossible to reach an agreement over the quality of the schema mapping. Therefore:

Result II. *The necessary condition for the existence of the Pareto-efficient equilibrium is that at least ϕ_0 percent of agents $p_i \in P$ conclude that $\gamma_i\,(1-\alpha_i) < (1-\phi_0)$.*

From a simple comparison of results I and II it is derived that the necessary condition for the existence of the Pareto-efficient equilibrium is reduced by a factor $(1-\alpha)$ in the developed strategy; how the sufficient condition to this equilibrium is provided by this strategy is discussed in the next section.

(ii) if $\phi_0 \leqslant \alpha \leqslant 1$, then in this case the probability distribution of θ is reduced to:

$$\pi^*(\theta) = \begin{cases} 0; & \theta = \theta_1 \\ 1; & \theta = \theta_2 \end{cases}$$

Obviously in this case, the Pareto-efficient equilibrium has been selected or, equivalently, semantic interoperability has emerged. Here, an agent can stop the adoption decision process (i.e. the result of the decision analysis would be certainly adoption), as a certain quality of schema mapping (ϕ_0) has already been achieved.

4.2. *Communication*

To benefit from the information that becomes available on the achieved consensus level in the advertisement decision model established in (5), the quality of the received partial mappings is taken into account for updating the probability of reaching the consensus, as already explained in Section 4.1. For risk formulation, we consider the snapshot of the estimated current quality of the schema mappings of other agents, as inferred based on the quality of the received sample of the partial mappings. Recall that the quality of the schema mappings at the other agents, as locally estimated by the received advertisements, is represented by the random variable X with probability distribution $f(x)$. Whereas the *quality* of the schema mapping of the agent community and the *consensus level* are equal in distribution, x can be considered to be independent from ϕ, because the received partial mappings are a random sample of the overall partial mappings. Thus, $f(x|\phi) = \pi(\phi)$. The Bayes risk of an advertisement decision δ, with respect to a prior distribution π on ϕ, is defined by an expected loss in [0,1] based on averaging over the random variable X [9] as follows:

$$r(\delta) = E^{\pi} E^{x}\big[L(\phi, \delta)\big] \tag{10}$$

where $L(\phi, \delta)$ is the loss function of advertisement according to the consensus level and can be developed from utility theory [9] as follows.

Let x_0 denote the quality of local schema mapping that results from using the instance of μ, which is subject to be advertised or not. We consider $L(.)$ to be in [0, 1] where, the choice of this interval serves only to set the scale for L. Given the assumption that the loss depends on the probability that an advertised mapping is adopted by other agents, the highest and the lowest losses are associated with δ_2 (advertising) and δ_1 (not advertising), respectively, when the relative quality of x_0 to x is low. More specifically, if $kx_0 < x$, then $L(\delta_2) = 1$ and vice versa $L(\delta_1) = 0$; where k is a tolerance factor that can be defined as a statistic metric on the resulting quality of the local schema mapping based on the received advertisements.

When $kx_0 > x$, the higher the quality of the instance is, the lower the probability that other agents will ignore the advertisement and the less the loss will be. Therefore, in this case, we assume $L(\delta_1) = x_0$ and $L(\delta_2) = 1 - x_0$. To this end, $L(.)$ is given by

$$L(\phi, \delta_1(x)) = \begin{cases} x_0; & x < kx_0 \\ 0; & x \geqslant kx_0, \end{cases}$$

$$L(\phi, \delta_2(x)) = \begin{cases} 1 - x_0; & x < kx_0 \\ 1; & x \geqslant kx_0 \end{cases}$$

To determine the decision variable δ, the risk of each advertisement rule δ_1 and δ_2 is assessed in accordance with Eq. (10) as follows:

$$r(\delta_1) = E_\phi\left[\int L(\phi, \delta_1(x))\, dF^X(x|\phi)\right] = \int_0^{kx_0} x_0\, dx = kx_0^2$$

$$r(\delta_2) = \int_0^{kx_0} (1 - x_0)\, dx + \int_{kx_0}^1 dx = 1 - kx_0^2$$

where $dF^X(x|\phi) = \pi(\phi)dx$. Therefore, by applying $r(\delta_1)$ and $r(\delta_2)$ in Eq. (5), the optimal advertisement rule δ^* is determined by individual agents:

$$\delta^* = \begin{cases} \delta_1; & x_0 \leqslant x^* \\ \delta_2; & x_0 > x^* \end{cases} \tag{11}$$

where $x^* = \{c_{adv}/((c_{adv} + c_a)\cdot k)\}^{1/2}$.

In summary, every autonomous service agent is guided by the developed strategy to decide whether to adopt or reject an advertisement, with respect to its local perception of the global consensus level; meanwhile, the local perception of $\pi(\phi)$ is estimated by taking advantage of the information that becomes available through local advertisements. In the next section, we investigate if following the strategy leads the system into the Pareto-efficient equilibrium and to the emergence of semantic interoperability.

4.3. Equilibrium selection

In the employed game models, we investigate the necessary conditions that should be held in order *adopt* – a received instance of a partial mapping μ – to be selected as a dominant strategy. However, in general, all choosing *reject* could also emerge as an equilibrium in the system. In the case that multiple equilibria exist, one way that one of them can finally emerge is to become a focal point on which the expectations of the players (participating agents) converge; this focal equilibrium could be determined by some process of preplay communication [15]. In the developed strategy, the agents can become focused on one of the equilibria through the communication mechanism proposed in Section 3.2.

As mentioned, in the proposed advertisement mechanism, a rational agent avoids advertising its instance of mapping μ when it has a high probability of being ignored by others. According to the definition[10] of α, it is reasonable to assume that in such a setting the expected value of α by individual agents gradually increases; hence, the decentralized strategy has the potential to focus the attention of agents on the Pareto-efficient equilibrium. We can also give a new interpretation to (9), stating the necessary condition for the existence of the Pareto-efficient equilibrium in the Bayesian game with respect to α. Specifically, the dominant strategy a^{π^*} for the agents is given by

$$a^{\pi^*} = \begin{cases} a_1; & \alpha \leqslant \alpha^* \\ a_2; & \alpha > \alpha^*, \end{cases} \tag{12}$$

where $\alpha^* = 1 - (1 - \phi_0)\gamma^{-1}$.

In other words, the necessary condition for the emergence of the Pareto-efficient equilibrium is that the assessment of α_i, based on the local observations of

[10] Recall that α is considered by individual agents, as the minimum quality of the schema mapping results from using the received advertisements on μ.

the participating agents $p_i \in P$ on the quality of the schema mapping, satisfies $\alpha_i > \alpha^*$. Therefore:

Result III. *The necessary and sufficient conditions for the Pareto-efficient equilibrium by the developed strategy are* $\gamma < (1-\phi_0)/(1-\alpha)$ *and that the sequential schema quality increments, denoted by* $\Delta\alpha$*, satisfy*

$$\Delta\alpha \geqslant \zeta\big(1-\alpha^*\big)(1-\gamma) \tag{13}$$

for some constant ζ *with* $0 < \zeta < 1$.

As analyzed in Section 4, $\gamma < (1-\phi_0)/(1-\alpha)$ is the necessary condition for the *adopt* strategy to be a Nash equilibrium of the game. However, in order for the consensus (i.e., Pareto-efficient equilibrium) to be reached, it should also be $\Delta\alpha \geqslant \zeta(1-\alpha^*)(1-\gamma)$. Based on the *geometric improvement* theorem from [3], it can be proved that in this case $\alpha \to \phi_0$; hence, the Pareto-efficient equilibrium is achieved (as discussed in Section 4.1, situation **(ii)**). Regarding the proof of Eq. (13), suppose that α^* is the initial value of α, α^i is the value of α at the ith iteration, and α^M the maximum value. In our setting, $\alpha^* = 1-(1-\phi_0)/\gamma$, and $\alpha^M = 1$. According to the geometric improvement theorem, if our developed strategy guarantees that for every iteration i, $\alpha^{i+1}-\alpha^i \geqslant \zeta(\alpha^M-\alpha^*)$ for some constant $0 < \zeta < 1$, then it terminates after at most $2\log(\alpha^M-\alpha^*)/\zeta$ iterations.

$$\begin{aligned}
\alpha^{i+1}-\alpha^i &\geqslant \zeta\big(\alpha^M-\alpha^*\big)\\
&= \zeta\left(\phi_0 - 1 + \frac{1-\phi_0}{\gamma}\right)\\
&\geqslant \zeta\big(\phi_0\big(1-\gamma^{-1}\big)-\big(1-\gamma^{-1}\big)\big)\\
&= \zeta(1-\phi_0)\big(\gamma^{-1}-1\big)\\
&= \zeta\left(\frac{1-\phi_0}{\gamma}\right)(1-\gamma)\\
&= \zeta\big(1-\alpha^*\big)(1-\gamma) \qquad \square
\end{aligned}$$

that is Eq. (13). When α tends to ϕ_0, the Pareto-efficient equilibrium becomes a focal point on which the expectations of the service agents converge. Result III is verified in the experimental results of Section 5.

The above statements are also supported by studying the trajectories of the strategies (adoption or rejection) followed by the agents with respect to the observed quality of the schema mapping. Considering that the successful strategy spreads among the agents with respect to the quality of the schema mapping through a proportional imitation rule,[11] we employ the replicator dynamics equation [32] for the game among homogeneous agents, which is given by

$$\frac{\dot{x_i}}{x_i} = V_i - \bar{V} \tag{14}$$

where x_i is the fraction of agents play strategy i, $\dot{x}$ is the derivative of x with respect to time, $V_i = E^{\pi^*}[U(a_i|\theta)]$ is the expected payoff of pure strategy i and $\bar{V}$ is the average expected payoff by all strategies. Assuming that a fraction x_1 of the agents play *reject* and $x_2 = 1-x_1$ of them play *adopt*, then the *per capita* increase of x_1 over time is given by

$$\begin{aligned}
\frac{\dot{x_1}}{x_1} &= \big(E^{\pi^*}\big[U(a_1|\theta)\big]-E^{\pi^*}\big[U(a_2|\theta)\big]\big)(1-x_1)\\
&= \big(c_a-[1-\phi_0]/[1-\alpha][b-c_h]\big)(1-x_1)
\end{aligned} \tag{15}$$

Adopt is an *evolutionary stable strategy*[12] (ESS) of the system when

$$\dot{x_1}/x_1 \leq 0 \;\Leftrightarrow\; \gamma(1-\alpha) < 1-\phi_0, \quad \forall x_1 \in [0,1] \tag{16}$$

In other words, if $\gamma(1-\alpha) < 1-\phi_0$ for the participants, then the system will asymptotically converge to the equilibrium where every agent adopts a received instance of an attribute mapping, which results in a more qualified schema mapping. For a system of heterogeneous agents, the system would still asymptotically evolve to the same equilibrium where every agent p_i plays *adopt* as long as $\gamma_i(1-\alpha_i) < 1-\phi_0$, as proved in the next section by simulation experiments. In this case, the consensus on the quality of the schema mapping is achieved if the expected value of α_i also increases gradually (as stated formally in (13)) during the adoption process.

[11] With a proportional imitation rule in a game, which effectively says: "imitate actions that perform better, with a probability proportional to the expected gain", we can use the usual replicator dynamics [22].

[12] An evolutionarily stable strategy [28] (ESS) is a strategy which, if adopted by a population of players, cannot be invaded by any alternative strategy that is initially rare.

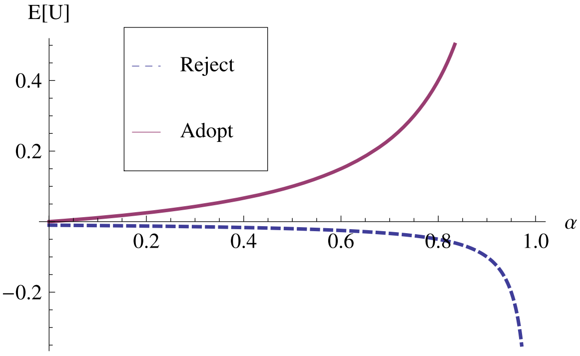

Fig. 3. Expected strategy payoffs of the Bayesian game for $b = 1$, $c_a = c_h = 0.1b$, $\phi_0 = 0.9$.

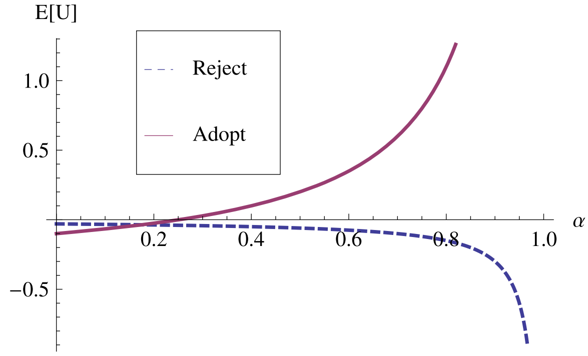

Fig. 4. Expected strategy payoffs of the Bayesian game for $b = 1$, $c_a = 0.4b$, $c_h = 0.1b$, $\phi_0 = 0.7$.

5. Evaluation results

In this section, we first evaluate the results of our analysis through a numerical example for better clarity. Then, we setup a set of experiments to verify how well the simulation tracks the predicted results of our theoretical analysis in different simulated settings.

5.1. Numerical example

We first consider a homogeneous system of agents with semantically diverse schema mappings, where the marginal adoption cost c_a is assumed to be 10% of the per-agent marginal benefit b, when the consensus is reached, the marginal cost of heterogeneity c_h equals the adoption cost and the consensus quality threshold ϕ_0 is considered to be high and equal to $\phi_0 = 0.9$ (see Section 3.1 for detailed description of the cost and benefit parameters). The strategy (adoption and rejection) payoffs are depicted in Fig. 3 for various values of α that is proxy of the consensus level reached. As the figure shows, adopting an advertisement that results in a more qualified schema mapping, is the dominant strategy for agents, if the cost of adoption is relatively low. Furthermore, the dominant strategy is determined regardless of the locally observed quality of the schema mappings, which results from using the received advertisements in this case. In this example, when the cost of adoption is higher than 10% of b, e.g., $c_a = 0.4b$ whereas still $c_h = 0.1b$, and the consensus quality threshold is lower than 0.9, e.g., $\phi_0 = 0.7$, as depicted in Fig. 4, the strategy payoffs would have a crosspoint at $\alpha^* \simeq 0.18$. Specifically, according to Eq. (12) and compatible to Eq. (16), with the mentioned values of cost and benefit, a quality greater than $a^* = 0.18$ for the schema mapping, relative to the specified consensus quality threshold, should be estimated locally by individual agents in order for the strategy *adopt* to have a payoff higher than the *reject* one.

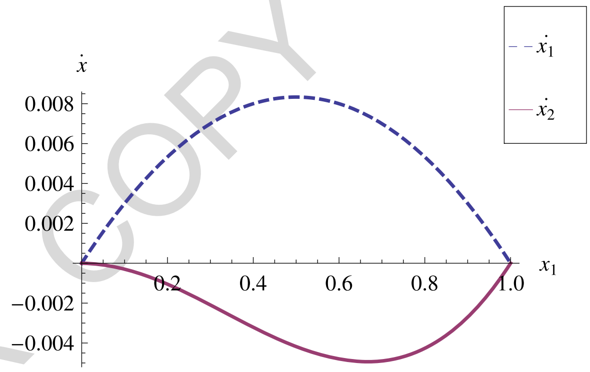

Fig. 5. Strategy evolutionary dynamics for $b = 1$, $c_a = 0.4b$, $c_h = 0.1b$, $\phi_0 = 0.7$ and $\alpha = 0.1$.

Next, we further analyze the effect of the locally-estimated level of consensus, based on the received advertisements, by means of evolutionary strategy dynamics. We consider the above homogeneous system with parameters $b = 1$, $c_a = 0.4b$, $c_h = 0.1b$, $\phi_0 = 0.7$, where a fraction x_1 of participating agents play *adopt* and the rest $x_2 = 1 - x_1$ play *reject*. When a low $\alpha = 0.1$ is locally-observed by the agents, the evolutionary dynamics $\dot{x_1}$ and $\dot{x_2}$ of the strategies with respect to x_1 are depicted in Fig. 5. As expected, if $\alpha < a^* = 0.175$, then the system asymptotically converges to the *reject* strategy, which means that the participating agents prefer to keep their semantically diverse schema mappings. However, as depicted by the evolutionary dynamics of the strategies in Fig. 6, for $\alpha = 0.2$, adopting the received advertisements that result in more qualified schema mapping is the evolutionary stable strategy in the system.

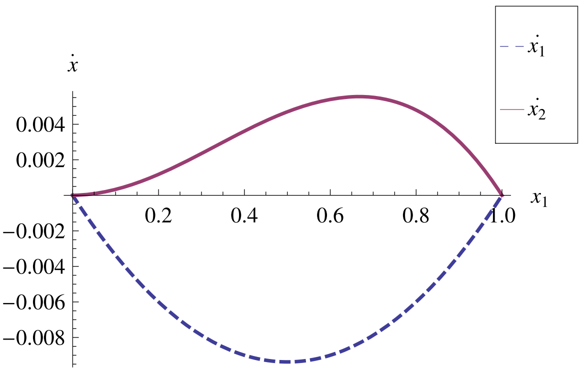

Fig. 6. Strategy evolutionary dynamics for $b = 1$, $c_a = 0.4b$, $c_h = 0.1b$, $\phi_0 = 0.7$ and $\alpha = 0.2$.

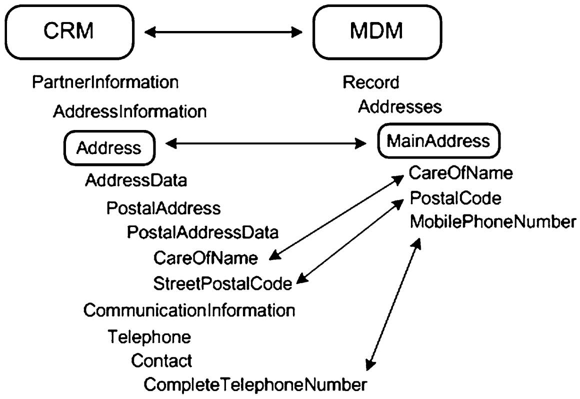

Fig. 7. A fragment of the mapping between CRM and MDM.

5.2. Simulation results

We simulate the running example in Section 2 with a set of $N = 40$ service agents registered in one registry that employs a given schema to collectively create a schema mapping with the schema of a second registry. The schema mapping method employed is attribute correspondences. As previously discussed, we can reasonably assume a market increase for those agents that reach consensus and a market decrease for the agents that stay out. Considering the importance of semantic agreement based on these assumptions, we would derive the cost and benefit of interoperability for the service agents in our simulation. In order to make the situation more realistic, we employ two real schemas describing business partners of SAP, namely customer relationship management (CRM) and master data management (MDM). A consensus threshold of $\phi_0 = 0.9$ is considered by the interacting agents for their collectively constructed schema mapping. We assume that the schema mapping consists of $A = 30$ attribute mappings (i.e. partial mappings), each of which is subject to advertisement by an agent. Without loss of generality, we assume that an advertised attribute mapping is made available for a limited amount of time by the registry to all the registered service providers. In our setting, to generate the mappings between the schemas of the two registries, the agents initially use AMC [30] and COMA++ [4] as two different mapping tools, and the mappings are given as XML path correspondences, e.g., *(schema1.path1 : schema2.path2)*. A fragment of the mapping between CRM and MDM is briefly illustrated in Fig. 7. Initially, in each experiment, individual agents autonomously generate an instance of the schema mapping by using the aforementioned mapping tools. Although an agent is not aware of the best qualified mapping, it can measure the relative quality of its current instance of the schema mapping, as follows. The quality metric that is used in both AMC and COMA++, is a combination of four matchers: (1) NAME that only considers the element names, (2) PATH that matches elements based on their hierarchical structure, (3) PARENTS, this structural matcher only considers the parent elements to estimate the similarity between two child elements, (4) LEAVES, this structural matcher only considers the leaf elements to estimate the similarity between two inner elements. The quality of an attribute correspondence is the weighted average of the similarity values given by these matchers. For example, the calculated quality for the mapping (CRM.PartnerInformation.AddressInformation. Address.AddressData.PostalAddress.PostalAddress Data.CareOfName : MDM.Record.Addresses. MainAddress.FaxCompleteNumber) is 0.45, and based on the aforementioned combinatorial metric, the mapping (CRM.PartnerInformation.AddressInformation. Address.AddressData.PostalAddress.PostalAddress Data.CareOfName : MDM.Record.Addresses. MainAddress.CareOfName) would obviously have a higher quality of 0.82. The output of AMC and COMA++ is a list of attribute correspondences and their similarity scores. Although both AMC and COMA++ use the same matching components (i.e. NAME, PATH, PARENTS, LEAVES), the overall quality (i.e confidence values) of these tools is different, because, for the different components, different heuristic functions are used. In our experimental setup, each agent calculates the overall quality of his local schema mapping as a weighted average of the scores of its individual attribute correspondences, whereas the weights are different for different agents. Therefore, the agents do not have to converge to the same schema mapping. However, note that this quality metric is monotonic to the score of any individual constituent attribute correspondence.

We perform each experiment in a number of rounds. In each round, each agent decides whether or not to take the risk of advertising an attribute mapping of its current schema mapping, according to the advertisements that it has seen in a specific time window of size w; the size of window can be specified in terms of a number of rounds or a number of received advertisements. Also, each agent decides on whether to adopt or reject a received advertisement on an attribute mapping, if using it results in a higher quality of its local schema mapping than it current has.

In each experiment, we evaluate the Precision [13] of the schema mapping instances of different service agents at each round. The best qualified overall schema mapping is defined as a perfect set of mappings, provided by an expert. Precision expresses the proportion of correct mappings in the schema mapping instance of an agent, as compared to the mappings provided by the expert. Hence, this metric can capture the relative quality of the current instance to the best qualified schema mapping. The average quality per round is illustrated to investigate the emergence rate of the desired semantic interoperability.

We perform all the experiments for homogeneous and heterogeneous agents (in terms of costs and benefit) as well. We illustrate the effect of changing each individual parameter of our model, while keeping the other parameters fixed. We present the results of the homogeneous and heterogeneous settings together for each experiment; the values of the parameters in the homogeneous setting are considered as the mean values of respective parameters in the corresponding heterogeneous setting.

The main purpose of the following simulation experiments is to investigate the emergence of semantic interoperability and the convergence rate of it. Hence, the cost and benefit parameters in the homogeneous setting or their mean value in the heterogeneous setting are set so that the necessary condition (previously derived through the performed analysis) for establishing the consensus is provided. However, we also demonstrate in the first experiment that if the necessary condition does not hold, there is no consensus on the quality of the schema mapping, as we expected.

First, we investigate the decision process of adoption in the set of autonomous agents. As the ratio of γ, which is defined in Eq. (4), is the determining parameter for the adoption decision of agents, we keep the advertisement cost c_{adv} constant and run the experiment for different values of γ. To investigate the convergence rate to emergence of consensus, the average quality (precision) per round is illustrated in Fig. 8. As we expected, if γ is greater than a specific threshold (e.g., for a higher adoption cost c_a), the necessary condition for establishing the consensus is not satisfied. In this case, each agent prefers to keep its current schema mapping instead of adopting more qualified partial mappings, hence the improvement in the average quality of the schema mapping is insignificant (e.g., see Fig. 8 for $\gamma = 0.3$). As experimentally found, the γ threshold is between 0.25 and 0.3, which is close to our analytical prediction. Also, the minimum quality α of the schema mapping is approximately 0.65, as depicted in Fig. 9. Thus, according to our analysis, and specifically Eq. (9), only a value of γ lower than 0.286 would result in the convergence of the schema quality to $\phi_0 = 0.9$ or, equivalently, in the emergence of se-

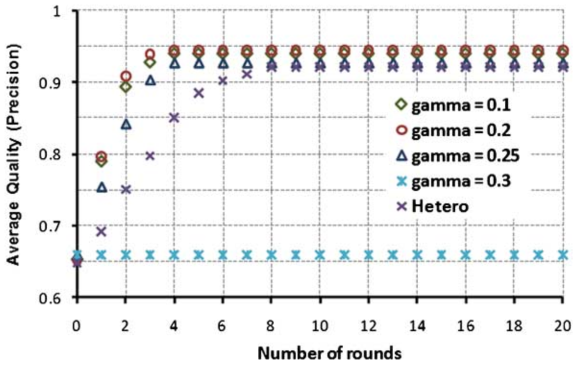

Fig. 8. The effect of increasing γ for $c_{adv} = 0.1c_a$, $A = 30$, $N = 40$, $\phi_0 = 0.9$, and in the heterogeneous setting, $\bar{\gamma} = 0.25$.

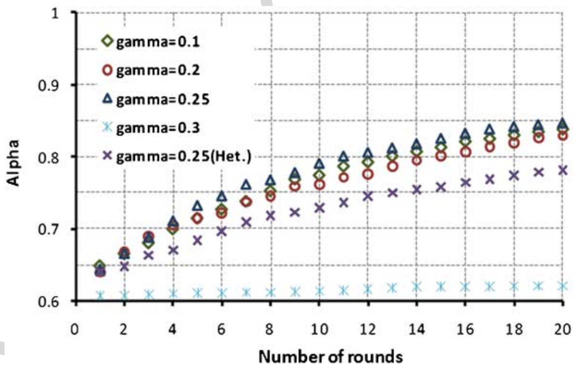

Fig. 9. The evolution of the expected value of α for $c_{adv} = 0.1c_a$, $A = 30$, $N = 40$, $\phi_0 = 0.9$, and in the heterogeneous setting, $\bar{\gamma} = 0.25$.

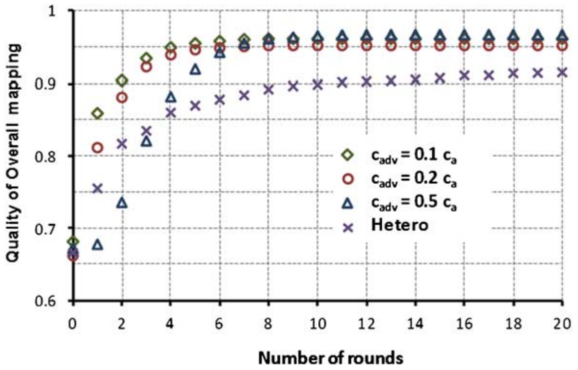

Fig. 10. The effect of increasing advertisement cost for $b = 1$, $c_a = c_h = 0.1b$, $A = 30$, $N = 40$, $\phi_0 = 0.9$, and in the heterogeneous setting, $\bar{c}_{adv} = 0.1\bar{c}_a$.

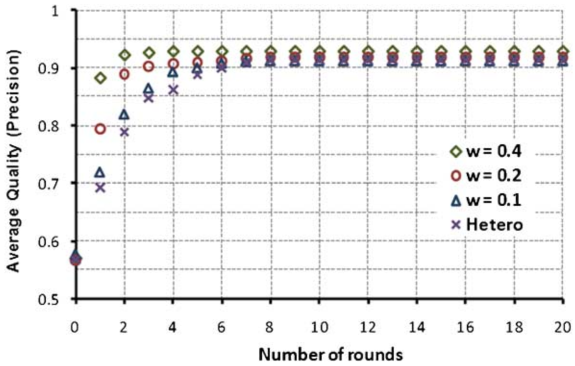

Fig. 11. The effect of increasing w for $b = 0.4$, $c_a = c_h = 0.25b$, $c_{adv} = 0.1c_a$, $A = 30$, $N = 40$, $\phi_0 = 0.9$, and in the heterogeneous setting, $\bar{w} = 0.3$.

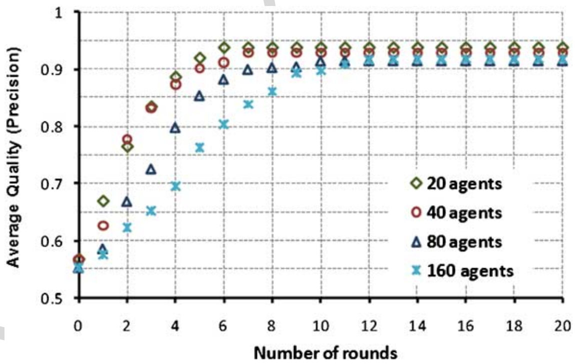

Fig. 12. The effect of increasing the number of participating agents for $b = 0.4$, $c_a = c_h = 0.25b$, $c_{adv} = 0.1c_a$, $A = 30$, $\phi_0 = 0.9$.

mantic interoperability. To scrutinize the results of this set of experiments, we also plot the evolution of the expected value of α for each experiment separately in Fig. 9. It is clearly illustrated in the figure that the sequential increments in the expected value of α satisfies the stated condition in Eq. (13) for low-enough values of γ. Therefore, the sufficient condition for reaching the Pareto-efficient equilibrium is satisfied for γ values lower than 0.286, as discussed in Section 4.3.

Next, we investigate the decision process of advertisement for the individual agents. To this end, as illustrated in Fig. 10, we keep the benefit, the adoption cost and the heterogeneity cost constant, and we vary the advertisement cost in different runs of the experiment. As experimentally shown (see Fig. 10), the higher the advertisement cost is, the longer the time is to reach the same average quality of the schema mapping, as a smaller number of agents decide to take the risk of advertising. It is clear that, although not illustrated, for a relatively high cost of advertisement, there would be no consensus on the quality of the schema mapping, because the agents would avoid advertising their partial mappings altogether, as we expected by Eq. (5).

We are also interested in seeing the effect of changing the parameter w on the convergence rate; recall that w is the size of a window in which the received advertisements are considered for estimating the global consensus level reached in the process of decision making according to Eq. (11). The size of a window is defined in terms of the number of service agents from whom an individual agent receives advertisements. In order to explicitly model the limited scope of advertisements through local communication of service agents, w is expressed as a fraction $\{0.4,0.2,0.1\}$ of N in these experiments. As depicted in Fig. 11, the establishment of consensus is guaranteed, even if a service agent only exchanges advertisements with a fraction of other agents. As the fraction of participating agents with whom a service agent communicates increases, the convergence rate increases as well, as expected.

Finally, we investigate the scalability of the proposed strategy by increasing the number N of participating service agents (Fig. 12) and the number A of constituent attributes of their schema (Fig. 13) in two different sets of experiments. It should be noted that in a realistic setting, the amount of information that becomes available to an agent through local advertisements is not necessarily proportional to the number of participating agents. Therefore, in these experiments, we decrease w when increasing N, so that the number of advertisements locally received by other agents remains constant. As demonstrated in Figs 12 and 13, semantic interoperability emerges with an acceptable convergence rate, even under this communication constraint, yet requires more rounds, as expected.

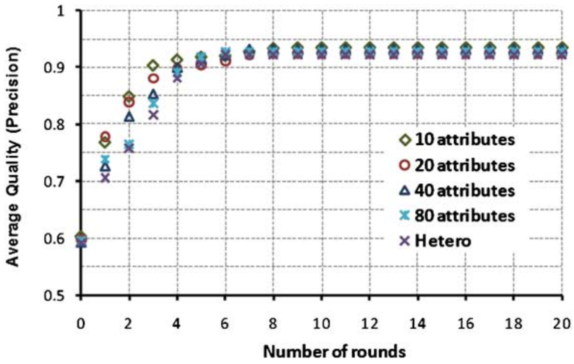

Fig. 13. The effect of increasing the number of attributes for $b = 0.4$, $c_a = c_h = 0.25b$, $c_{adv} = 0.1c_a$, $N = 40$, $\phi_0 = 0.9$, and in the heterogeneous setting, $A = 80$.

6. Related work

Due to the importance of interoperability as a necessary requirement of many real-world applications of multi-agent systems, this generic problem has been studied in various contexts. For example, some studies [29,33] focus on *norm emergence*, which addresses our problem in a very general form, as they study a type of multi-agent agreement process in which agents converge to a common strategy in a variety of situations.

In a more specific form, the problem of semantic agreement has also been thoroughly studied. A number of studies deal with semantic interoperability through the evolution of a common ontology or a global schema based on the cooperation of distributed participants [31,35,37]. However, the assumption of a common ontology is often too strong or unrealistic in cooperative multi-agent systems, as argued by [2]. There exist different approaches that enable the agents to keep their ontologies while improving their interoperability. For example to support the communication among heterogeneous agents, the authors in [11] propose an approach in which the agents gradually build towards a semantically integrated system by exchanging their individual ontological information on an as-need basis. Another approach is proposed by [2], in which the agents use machine-learning techniques to learn new concepts from other agents through instance examples to improve their communication - hence - cooperation abilities. A more practical approach is presented in [34] for the cases where there is not a sufficient number of instances shared by the agents. Their distributed approach for the creation and the dissemination of new concepts allows individual ontologies of the agents to evolve cooperatively, but according to the results of their own queries. Similarly, in our proposed approach for improving the interoperability of the agents, we avoid the assumption of common ontology. By taking this approach, whereas the agents are able to keep their individual schema or ontology by making use of the mappings between different schemas or ontologies, these mappings are not necessarily the same and their quality is improved by the rational participation of every agent.

According to [1], in order to reach agreements on common interpretations, the state of semantic interoperability as an emergent phenomenon depends on the efficiency of the conducted negotiations. Thus, the main concern in other studies is how to develop negotiation mechanisms, in order to establish a consensus on ontology/schema mappings used by distributed autonomous participants [26,27]. For example, the authors in [27] and [26] propose to use value-based argumentation frameworks [8] to deal with conducting negotiations between distributed agents who use different ontologies, in order to reach an agreement over ontology mappings. This is achieved by using an argumentation process in which candidate mappings are accepted or rejected, based on the ontological knowledge and the agents preferences. As opposed to these approaches, we take into account the rationality of agents in the emergence of the consensus on the schema mappings. Such rationality is naturally prevalent in the collaboration of independent business entities and has to be dealt with.

Another line of research employs distributed coordination for achieving semantic interoperability. For example, the authors in [36] propose setting up a semiotic dynamics that achieves this coordination. The agents self-organize an interlingua with labels whose semantics (i.e. the meanings of their underlying categories) are coordinated among the agents without central coordination. Whereas, in our case, not all the agents necessarily converge to the same semantic mappings, yet they can reach an agreement by establishing a consensus on quality of the schema mapping. In [7], the authors analyze a dynamic model known as the Naming Game and discuss that it is able to account for the distributed coordination of autonomous agents and the emergence of global agreement. Aiming at defining the microscopic behavior of the agents in emergence of shared vocabularies, the authors in [6] study semiotic dynamics of a similar model to [7]. They show that the model exhibits the same phenomena as observed in human semiotic dynamics, specifically a period of preparation followed by a rather sharp transition into global coordination. These developed models, however, cannot be applied in the context of establishing consensus on the quality of a schema mapping that we deal with in the present paper, because they involve two-way pairwise interactions, a different mapping quality verification method, and a different end of game.In a more generic setting, the authors in [19] address the problem of coordination in multi-agent systems for Collective Iterative Allocation. Using various individual agents proposals which are based on their estimations of the performance of different teams, a team is selected for a given task. The selection is based on group decision policies, where each agent contributes (e.g., through proposing) to the decision as to which team should be selected for a given task. After a team is selected and executes the task, they progressively develop a better understanding of the true performance of teams which is taken into account in subsequent allocations. As one of the important directions for future improvement of the algorithm, the authors have mentioned that they require an evaluation of the communication requirement of the agents. In our developed decentralized strategy of coordination, we have addressed this issue by taking into account the communication requirement of the agents.

Overall, whereas most of the aforementioned approaches focus on developing different techniques in order to establish interoperability, we shift our attention to investigate under which conditions semantic interoperability emerges. Strictly speaking, we take into account the situations in which agents have different costs and benefits for being involved in the establishment of the consensus on schema/ontology mappings and each agent tries to maximize its own utility. In the context of collaborative data sharing, one of the studies [38] has addressed semantic interoperability in a very similar setting to the sample scenario of our work. Although, the proposed decision making approach in [38] is predicated on *static individual-based* policies for the participants (to adopt or reject updates), whereas we develop a *rational* and *dynamic* decision making process for achieving consensus on the schema mapping qualities among the participating agents.

Finally, the game which is modeled and analyzed in this paper is closely related with the social dilemma occurring in public good or critical mass situations, in the literature of economic sciences. Although the qualitative dynamics of public good or critical mass games have been fully analyzed [21,25], they have some shortcomings as potential solutions for the emergence of semantic interoperability. As an important instance, they are not concerned with the heterogeneous motivations of individual participants (in terms of their cost and benefit) for being involved in the emergence process of establishing interoperability. Also, [18,20] aim to improve the odds for the implementation of the public good through some form of commitment (which is costly when it is fake) or deposit on behalf of the agents prior to the public good game. The purpose of that work is to discourage various free-riding strategies either by avoiding playing the game with free-riders or by restricting access to the public good for free-riders. As compared to [18,20], we do not consider prior commitment of agents for contributing to the public good, as such a setting often requires monetary transfers among agents. However, some sort of prior expression of interest on behalf of the agents would be indeed needed in reality for initiating the whole process of reaching semantic agreement. Note that, in our case, non-contributors (i.e., those that do not adopt or advertise partial mappings) are self-restricted from accessing the public good, i.e., they cannot free-ride, because each agent constructs the public good locally through some costly effort, as opposed to the work in [18,20].

7. Conclusion

In this paper, we have employed a game-theoretic approach to analyze the dynamics of bottom-up (i.e. through agent interaction) emergence of semantic interoperability in a distributed setting of rational agents. We analytically found that semantic emergence can arise, even as an evolutionary stable equilibrium, if certain conditions hold regarding the communication cost, the cost of technological adaptation, the expected mutual benefit from interoperability, as well as the expected loss from isolation. Our analysis significantly improves the understanding of the rational dynamics involved in an agreement bottom-up emergence process and explains that interoperability, in general, apart from being a challenging technical problem, also has economic implications in order to emerge. Our simulation experiments verify our analysis and demonstrate the emergence of semantic interoperability in realistic settings. Note that we have made some simplifying assumptions, such as the linear relationship between the quality of a schema mapping and the consensus, or considering an average benefit and cost for all agents in our evolutionary dynamics analysis. Although these assumptions might avoid some difficulties of real life, they enable us to employ more enriched analytical tools for scrutinizing a simple formal model in a domain that needs more attention and to gain insights that can be applied to the more complex situations. For future work, we plan to study different cases of bottom-up semantic interoperability emergence where limited communication among agents employing different schemas is possible, e.g., storage clouds.

References

[1] K. Aberer, P. Cudré-Mauroux, A. Catarci, M.-S. Hacid, A. Illarramendi, V. Kashyap, M. Mecella, E. Mena, E. Neuhold, O. Troyer, T. Risse, M. Scannapieco, F. Saltor, L. Santis, S. Spaccapietra, S. Staab and R. Studer, Emergent semantics principles and issues, in: *DASFAA'04*, 2004.

[2] M. Afsharchi, B.H. Far and J. Denzinger, Ontology-guided learning to improve communication between groups of agents, in: *AAMAS'06*, 2006.

[3] R.K. Ahuja, T.L. Magnanti and J. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, 1993.

[4] D. Aumueller, H. Do, S. Massmann and E. Rahm, Schema and ontology matching with coma++, in: *Proc. of the ACM SIGMOD International Conference on Management of Data*, ACM Press, NY, USA, 2005, pp. 906–908.

[5] M. Baldoni, C. Baroglio, A.K. Chopra, N. Desai, V. Patti and M.P. Singh, Choice, interoperability, and conformance in interaction protocols and service choreographies, in: *AAMAS'09*, 2009.

[6] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto and L. Steels, Sharp transition towards shared vocabularies in multi-agent systems, *Journal of Statistical Mechanics: Theory and Experiment*.

[7] A. Baronchelli, V. Loreto and L. Steels, In-depth analysis of the naming game dynamics: The homogeneous mixing case, *International Journal of Modern Physics C* **19**(5) (2008), 785–812.

[8] T. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, *Journal of Logic and Computation* **13** (2003), 429–448.

[9] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

[10] D.G.I. for Standardization, *The Economic Benefits of Standardization*, Beuth, Verlag, 2008.

[11] J. Diggelen, R. Beun, F. Dignum, R. Eijk and J. Meyer, Anemone: An effective minimal ontology negotiation environment, in: *AAMAS*, 2006, pp. 899–906.

[12] A. Dixit and R. Pindyck, *Investment Under Uncertainty*, Princeton University Press, Princeton, 1994.

[13] F. Duchateau, Z. Bellahsène and E. Hunt, Xbenchmatch: A benchmark for XML schema matching tools, in: *Proc. of the 33rd International Conference on Very Large Data Bases (VLDB'07)*, Vienna, Austria, 2007, pp. 1318–1321.

[14] ebXML, from http://www.ibm.com/developerworks/xml/library/x-ebxml/.

[15] T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design*, Prentice Hall, 2005.

[16] J. Euzenat and P. Shvaiko, *Ontology Matching*, Springer, 2007.

[17] A. Gal, *Uncertain Schema Matching, Synthesis Lectures on Data Management*, Morgan & Claypool Publishers, 2011.

[18] A. Gerber and P.C. Wichardt, Providing public goods in the absence of strong institutions, *Journal of Public Economics* **93**(3–4) (2009), 429–439.

[19] C. Guttmann, M. Georgeff and I. Rahwan, Collective iterative allocation: Enabling fast and optimal group decision making, the role of group knowledge, optimism, and decision policies in distributed coordination, *Web Intelligence and Agent Systems: An International Journal* **8**(1) (2010), 1–35.

[20] T.A. Han, L.M. Pereira and T. Lenaerts, Avoiding or restricting defectors in public goods games?, *Journal of the Royal Society Interface* **12** (103) (2014).

[21] C. Hauert, N. Haiden and K. Sigmund, The dynamics of public goods, *Discrete and Continuous Dynamical Systems, Series B* **4**(3) (2004), 575–587.

[22] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, 1998.

[23] Z. Ives, T. Green, G. Karvounarakis, N. Taylor, P.T.V. Tannen, M. Jacob and F. Pereira, The orchestra collaborative data sharing system, *ACM SIGMOD Record* **37**(3) (2008), 26–32.

[24] J.J. Jung, Service chain-based business alliance formation in service-oriented architecture, *Expert Systems with Applications* **38**(3) (2011), 2206–2211.

[25] C. Keser, I. Suleymanova and C. Wey, Technology adoption in critical mass games: Theory and experimental evidence, DIW Berlin Discussion Paper, No. 961.

[26] L. Laera, I. Blacoe, V. Tamma, T. Payne, J. Euzenat and T. Bench-Capon, Argumentation over ontology correspondences in MAS, in: *AAMAS'07*, 2007, pp. 1–8.

[27] L. Laera, V. Tamma, J. Euzenat, T. Bench-Capon and T. Payne, Reaching agreement over ontology alignments, in: *ISWC'06*, 2006.

[28] S.J. Maynard, *Evolution and the Theory of Games*, Cambridge University Press, 1982.

[29] P. Mukherjee, S. Sen and S. Airiau, Norm emergence under constrained interactions in diverse societies, in: *AAMAS'08*, 2008.

[30] E. Peukert, J. Eberius and E. Rahm, AMC: A framework for modelling and comparing matching systems as matching processes, in: *Proc. International Conference on Data Engineering*, Hanover, Germany, 2011.

[31] P. Plessers, O. De Troyer and S. Casteleyn, Understanding ontology evolution: A change detection approach, *Web Semantics* **5**(1) (2007), 39–49.

[32] L. Samuelson, *Evolutionary Games and Equilibrium Selection*, MIT Press Books, The MIT Press, 1998.

[33] S. Sen and S. Airiau, Emergence of norms through social learning, in: *IJCAI'07*, 2007.

[34] M. Şensoy and P. Yolum, A cooperation-based approach for evolution of service ontologies, in: *AAMAS'08*, 2008.

[35] K. Siorpaes and M. Hepp, myOntology: The marriage of ontology engineering and collective intelligence, in: *Bridging the Gep Between Semantic Web and Web 2.0 (SemNet 2007)*, 2007.

[36] L. Steels and P. Hanappe, Interoperability through emergent semantics a semiotic dynamics approach, *Journal on Data Semantics* **VI** (2006), 143–167.

[37] V. Tanasescu and O. Streibel, Extreme tagging: Emergent semantics through the tagging of tags, in: *International Workshop on Emergent Semantics and Ontology Evolution*, 2007.
[38] N. Taylor and Z. Ives, Reconciling while tolerating disagreement in collaborative data sharing, in: *Proc. of ACM SIGMOD International Conference on Management of Data*, 2006, pp. 13–24.
[39] G. Vakili, T.G. Papaioannou and K. Aberer, Analyzing the emergence of semantic agreement among rational agents, in: *13th IEEE Conference on Commerce and Enterprise Computing*, CEC, Luxembourg, 2011, pp. 161–166.
[40] G. Vetere and M. Lenzerini, Models for semantic interoperability in service-oriented architectures, *IBM Systems Journal* **44**(4) (2005), 887–903.
[41] M. Wilkinson, Interoperability with Moby 1.0 – it's better than sharing your toothbrush!, *Briefings in Bioinformatics* **9**(3) (2008), 220–231.