# Analyzing the Emergence of Semantic Agreement among Rational Agents

Golnaz Vakili[1], Thanasis G. Papaioannou[2], Zoltán Miklós[2], Karl Aberer[2], Siavash Khorsandi[1]

[1]Dept. of Computer Engineering and Information Technology
Amirkabir University of Technology, Tehran, Iran
Email: {g_vakili, khorsandi}@aut.ac.ir

[2] School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Email: firstname.lastname@epfl.ch

***Abstract*—Todays complex online applications often require the interaction of multiple services that potentially belong to different business entities. Interoperability is a core element of such an environment, yet not a straightforward one. In this paper, we argue that the emergence of interoperability is an economic process among rational agents and, although interoperability can be mutually beneficial for the involved parties, it is also costly and may fail to emerge. As a sample scenario, we consider the emergence of semantic interoperability among rational service agents in the service-oriented architectures (SOA) and 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.**

***Index Terms*—schema mappings; incentives; game theory; risk; evolutionary stability**

## I. Introduction

Todays emerging complex online applications often require the orchestration of multiple services that potentially belong to different business entities. According to the service-oriented architecture (SOA) design principles, services should be flexible to discover, select, and use other services to fulfill a given task and goal [1]. Service interaction (i.e. communication) can be achieved by syntactic interoperability, while conformance to the service roles in the service orchestration is pertained to semantic interoperability [2]. Standardization efforts have a long history dealing with syntactic interoperability, while semantic interoperability is still open and mainly dealt with by ontology alignment, schema mapping and attribute correspondence approaches [3]. However, apart from the technical challenges for achieving interoperability, the latter is often a bottom-up dynamic decision making process among self-interested industrial agents that involves costly information exchange and technological adaptation for conformance to the agreed standards [4]. The benefits from adopting standards [5] as well as the decision making process regarding the optimal timing for the adoption of a new standard [6] have 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 [funded by the EU project NisB (FP7-ICT 256955)], we employ a game-theoretic approach to analyze the individual incentives of rational agents in decision-making towards the emergence of semantic agreement. As a running example, we consider the real problem of achieving semantic interoperability among rational service agents in the SOA context that employ different schemas or ontologies to describe their interfaces. Service agents that employ a common schema and wish to utilize services from another agent that employs a different schema have to collectively reach an agreement on their local semantic mapping of the former schema to the later one through costly iterative communication and mapping re-adjustment. However, if semantic agreement is reached, a benefit is expected to be mutually enjoyed by the service agents involved in the semantic mapping creation process, while 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. We identify as 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) have been identified and verified by simulation experiments. Our modeling can be applied in more general distributed settings for achieving interoperability among rational agents.

## II. 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 motivation of this sample scenario is to address similar semantic interoperability problems encountered in real applications of web services and other settings. We consider a system of distributed service repositories where service providers register their service interfaces. At each registry, service descriptions are represented based on a specific structured schema. To keep our analysis generic, we do not make any assumption on the specific formalism of schemas. For example, they could be ontologies or XML schemas. Interoperability between any two sets of service providers registered in two different registries requires the existence of *semantic mappings* between their corresponding schemas. By semantic interoperability, service agents at the first registry can employ the services of the agents at the second registry and vice versa. E.g. 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 to constitute a value-added social-event ticketing service. Thus, semantic interoperability increases the business potential for the service agents at either registry. Any subset of the service providers that are registered at the same registry may communicate to each other. The formalism of such schema mappings is beyond the scope of our work and, without loss of generality, simple attribute correspondences can be assumed as constituents. Semantic interoperability in this setting is established by a consensus among the set of service providers registered at one registry, on *acceptable quality* of the schema mapping that is going to be used in their collaboration with another set in the system. The more agents that reach a schema mapping of high quality, the higher the business potential for collectively constructing complex value-added services with service agents from the second registry.

Towards the schema mapping creation, initially, each agent at the first registry has a local version of the schema mapping, potentially constructed using commercial tools, such as AMC [7], COMA++ [8], etc, which have limited effectiveness. To this end, partial mappings have to be exchanged (through costly communication) among the agents of the same registry in order for the quality of the schema mapping to gradually improve towards a minimum acceptable level (which may be different for each agent), as the *ground truth* is unknown. Upon receiving a partial mapping each agent is able (using one of the quality metrics in [9]) to evaluate whether replacing its corresponding local partial mapping leads to an improvement in the schema mapping quality. However, 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, 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.

## III. Decentralized Strategy

A decentralized strategy is developed to coordinate rational decisions of the agents $P$ for collectively constructing a schema mapping of acceptable quality based on local communications.We refer to the proportion of the agents $P$ who reach agreement on a certain quality of a schema mapping to 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 a 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 $\Theta$ that is:

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

where $\phi_0$ is the *consensus level threshold* that is sufficient for the *agreement*. Therefore, we define that a consensus on quality of the schema mapping is established when at least $\phi_0$ proportion of the service agents $P$ reach an agreement on it. The decentralized strategy will guide the decisions of the agents on advertising their partial mappings on one hand, and adopting the received advertisements on the other hand.

### A. Adoption

The received instance of a mapping $\mu$ is taken into consideration by a rational agent if it results in a more qualified schema mapping in comparison with the currently used one (by employing one of the quality metrics in [9], e.g. average string similarity); otherwise, it will be ignored. To guide rational decisions of the agents, the decentralized strategy is developed based on a decision model which is constructed in accordance with the expected individual utility maximization. Each agent has two choices on receiving an instance of mapping $\mu$ which 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. The possible states of the uncertain parameter is $\theta_1 : no\,consensus$ and $\theta_2 : consensus$. Then, an agent will gain a *utility* $U(a|\theta)$ according to its action $a$ and depending on the state of $\theta$.

Semantic interoperability is 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 the received instances of $\mu$ if a consensus is established. Meanwhile, a costly effort $c_a$ is exerted by the agent for the eventual implementation and system integration of the partial mapping $\mu$. Thus, if an agent adopts the received instance of $\mu$ 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, yet without getting any benefit from it. On the other hand, if an agent uses a low quality schema mapping while others have collectively found one of acceptable quality, then he should pay a *heterogeneity cost*; this cost results from his low expected benefit from interoperability and his expected market losses from his semantic isolation. We denote $c_h$ as the marginal cost of heterogeneity that is expected to be incurred. Last, in case that an agent rejects a received instance of $\mu$ while there is eventually no consensus, then no cost or benefit occurs and thus no utility is obtained. To summarize:

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

This decision making process can be modeled as a positive-sum game among rational agents with respect to their derived utilities. Based on Nash equilibrium analysis, there will be 2 equilibria 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 $p_j \in P, p_j \neq p_i$. While the first equilibrium is Pareto-efficient and convinces semantic interoperability, this question arises that which equilibrium will be chosen by the system. It depends on the state of $\theta$ and actually the probability distribution of $\pi(\theta)$ that should be obtained. It can be concluded from maximum entropy priority [10] that $\pi(\phi) = U[\alpha, \beta]$ where $\alpha = 0$ and $\beta = 1$ to model a prior distribution for $\phi$ as non-informative as possible. Then, it is straightforward to assess the probability distribution of $\theta$ based on $\pi(\phi)$ with reference to (1). According to Bayes decision principle [10], a *rational* agent's decision maximizes $\rho(a, \pi)$ that $\rho$ is the Bayesian expected utility and is defined as:

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

Therefore, to make an optimal decision the expected utility of $a_1$ and $a_2$ is assessed respectively by each agent:

$$E^{\pi}[U(a_1|\theta)] = -c_h\,(1 - \phi_0), \quad E^{\pi}[U(a_2|\theta)] = b\,(1 - \phi_0) - c_a$$

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

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

Thus, when a received instance of mapping $\mu$ is taken under consideration by an agent $p_i \in P$, it would be *rejected* if $\gamma_i > (1 - \phi_0)$ and *adopted* if $\gamma_i < (1 - \phi_0)$. On the other hand, if the decision analysis results in $a_1$ as the Bayes action in more than $(1 - \phi_0)$ percent of agents, then it would be impossible to reach an agreement over the quality of the schema mapping; in this case with reference to (1), the proportion of the agents that prefer to keep and use their current instances of the schema mapping is too high to achieve a consensus on the quality:

**Result I.** *The necessary condition for existence of the Pareto-efficient equilibrium in the system is that at least $\phi_0$ 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 $\phi_0$ percent[1] of the agents $p_i \in P$, the $\gamma_i$ (the ratio of adoption cost to the cost and benefit of interoperability) is less than $1 - \phi_0$.

### B. Communication

We develop the advertisement decision making model of the decentralized strategy based on individual risk minimization. There is always a risk in advertisement of a partial mapping, as it will be ignored by the receiver agents if it does not result in a more qualified schema mapping. Thus, an agent should make a loss due to taking the risk of advertisement and incurring a cost of value $c_{adv}$ for the required communication. However, this should be compared with the *opportunity* loss of not advertising a given mapping, that comes by running the risk of not reaching a consensus because of not advertising a partial mapping of high quality (as compared to the quality of the received advertisements). (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.) In other words, if a consensus is expected to be reached, then the advertisement cost is expected to be compensated as 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, if the opportunity loss from not advertising a certain partial mapping is greater than the expected investment loss due to the advertisement, then the agent should advertise it. Formally, assuming a decision space $\Delta = \{\delta_2, \delta_1\}$ for advertisement, where $\delta_2$ and $\delta_1$ respectively corresponds to advertising or not a given partial mapping $\mu$, the agent chooses $\delta_2$ if the following condition is true:

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

where $c_{adv}$ is the advertisement cost, $c_a$ is the adoption cost, and $r(.)$ is the risk function. For risk formulation, 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 $\phi$. According to the definition of the *quality* and the *consensus level* of the schema mapping it is reasonable to assume that they are *equal in distribution*: $\forall x \in [0, 1]$, $P(Q \leq x) = P(\Phi \leq x)$. It can be argued in justification of this assumption that, the higher the quality of the schema mapping, the higher consensus level of it is expected (as mentioned, only a higher quality instance of a received partial mapping maybe adopted by a rational agent). As the *quality* of the schema mapping of the agent community and the *consensus level* are equal in distribution, while $x$ can be considered to be independent from $\phi$, thus, $f(x|\phi) = \pi(\phi)$. The Bayes risk of an advertisement decision is defined by [10]:

$$r(\delta) = E^{\pi} E^{x}[L(\phi, \delta)]\,, \quad (6)$$

where $L(\phi, \delta)$ is the loss function of advertisement according to the consensus level and can be developed from utility theory [10]. Let $x_0$ denote the quality of local schema mapping that results from using the instance of $\mu$, which is subject to be advertised. 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 to $\delta_2$ (advertising) and $\delta_1$ respectively when the relative quality of $x_0$ to $x$ is low. More specifically, if $kx_0 \leq x$, then $L(\delta_2) = 1$ and $L(\delta_1) = 0$; $k$ is a tolerance factor that can be defined as a statistic metric on the

[1] where threshold of $\phi_0$ can be defined to have the same value for all the participating agents or it can be set by individual agents to different values. We postpone investigation of the later setting to Section VI.

resulting quality of the local schema mapping based on the received advertisements. On the other hand, when $kx_0 > x$, the higher the quality of the instance, the lower the probability that other agents ignore the advertisement and the less the loss. In this case, we assume $L(\delta_1) = x_0$ and $L(\delta_2) = 1-x_0$. Then:

$$L(\phi, \delta_1(x)) = \begin{cases} x_0 & ; x < kx_0 \\ 0 & ; x \geq kx_0 \end{cases}, \; L(\phi, \delta_2(x)) = \begin{cases} 1 - x_0 & ; x < kx_0 \\ 1 & ; x \geq kx_0 \end{cases}$$

Substituting loss formulas to the equation (6) and applying $r(\delta_1)$ and $r(\delta_2)$ in equation (5), we obtain after some algebra the optimal advertisement rule $\delta^*$ determined by individual agents: $\delta^* = \delta_1$ if $x_0 < x^*$ and $\delta^* = \delta_2$ if $x_0 > x^*$; where $x^* = \{c_{adv}/((c_{adv} + c_a)\cdot k)\}^{1/2}$.

## IV. Bayesian Analysis of Decentralized Strategy

To take advantage of the information that becomes available through advertisements, the distribution of $\phi$ at the time of decision making should be the posterior distribution, $\pi(\phi|x)$. Then, it 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 $\phi$ over 0 and 1 is revised to model $\pi(\phi|x)$. According to the interest of the agents in reaching a consensus, the upper bound of $\pi(\phi|x)$ is always set to 1. The lower bound is set to $\alpha$, where based on the equality of $\Phi$ and $Q$ in distribution, $\alpha$ can be considered by individual agents, as the minimum quality of the schema mapping results from using the received advertisements. Thus: $\pi^*(\theta) = (\phi_0 - \alpha)/(1-\alpha)$ if $\theta = \theta_1$ and $(1-\phi_0)/(1-\alpha)$ if $\theta = \theta_2$. where $\pi(\theta|x)$ is denoted by $\pi^*(\theta)$. Based on the value of $\alpha$, two situations can be distinguished:

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

$$\begin{aligned} E^{\pi^*}[U(a_1|\theta)] &= -c_h\,(1-\phi_0)/(1-\alpha) \\ E^{\pi^*}[U(a_2|\theta)] &= b\,(1-\phi_0)/(1-\alpha) - c_a \end{aligned} \quad (7)$$

To determine the Bayes action $a^{\pi^*}$, we should find the crosspoint of these *monotonic* functions. After some algebra:

$$(1-\alpha)\gamma = (1-\phi_0) \quad (8)$$

Thus, in this situation, when a received instance of a partial mapping is taken under consideration by a rational agent $p_i \in P$, it would be *rejected* if $\gamma_i\,(1-\alpha_i) > (1-\phi_0)$and *adopted* if $\gamma_i\,(1-\alpha_i) < (1-\phi_0)$. On the other hand, as we 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 would be impossible to reach an agreement over the quality of the schema mapping:

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

It is concluded from results I and II that, the necessary condition for existence of the Pareto-efficient equilibrium is reduced by factor $(1-\alpha)$ in the developed strategy.

**Case (ii)** if $\phi_0 \leq \alpha \leq 1$, then in this case the probability distribution of $\theta$ is reduced to: $\pi^*(\theta) = 0$ if $\theta = \theta_1$ and $\pi^*(\theta) = 1$ if $\theta = \theta_2$. Obviously in this case, the Pareto-efficient equilibrium has been selected or equivalently semantic interoperability has emerged. Thus, an agent can stop the adoption process in this situation since a certain quality of schema mapping ($\phi_0$) has already been achieved.

## V. Equilibrium Selection

In the employed game models, we investigated the necessary conditions that should be held in order *adopt* - a received instance of a partial mapping $\mu$ - to be selected as a dominant strategy. However, in general, all choosing *reject* could also emerge as an equilibrium in the system. In case that multiple equilibria exist, one way that one of them may finally emerge is to become a focal point on which the expectations of the players converge; this focal equilibrium could be determined by some process of pre-play communication [11]. The proposed advertisement mechanism can be considered as such a pre-play communication in the developed strategy. In this mechanism, an agent avoids to advertise its mapping $\mu$ when it has a high probability to be ignored by others. It is reasonable then to assume that the expected value of $\alpha$ by individual agents gradually increases and 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 (8), stating the necessary condition for existence of the Pareto-efficient equilibrium in the Bayesian game with respect to $\alpha$. Specifically, the dominant strategy is:

$$a^{\pi^*} = \begin{cases} a_1 & ; \alpha < \alpha^* \\ a_2 & ; \alpha > \alpha^* \end{cases}, \quad (9)$$

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

The sufficient condition for the Pareto-efficient equilibrium is also provided by the developed strategy if the sequential increments in expected value of $\alpha$, denoted by $\Delta\alpha$ satisfy (as also demonstrated in Section VI):

$$\Delta\alpha \geq \zeta(1-\alpha^*)(1-\gamma) \quad (10)$$

for some constant $\zeta$ with $0 < \zeta < 1$. By using the *geometric improvement* theorem from [12], it can be proved that in this case $\alpha \to \phi_0$ and hence the Pareto-efficient equilibrium is achieved. In other words, when $\alpha$ tends to $\phi_0$ , the Pareto-efficient equilibrium becomes a focal point on which the expectations of the agents converge.

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. To this end, we employ the replicator dynamics equation for the game among homogeneous agents, which is:

$$\frac{\dot{x}_i}{x_i} = V_i - \bar{V}, \quad (11)$$

where $x_i$ is the fraction of agents plays 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 plays reject and $x_2 = 1 - x_1$ of them plays adopt, then the *per capita* increase of $x_1$ over time is given by:

$$\frac{\dot{x_1}}{x_1} = (c_a - [1 - \phi_0]/[1 - \alpha][b - c_h])(1 - x_1) \qquad (12)$$

If $\gamma(1 - \alpha) < 1 - \phi_0$ for the agents, then $\dot{x_1}/x_1 \leq 0$, $\forall x_1 \in [0, 1]$, and thus adopt is an *evolutionary stable strategy* of the system. In other words, if $\gamma(1-\alpha) < 1-\phi_0$for the participants, 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 $\gamma_i(1 - \alpha_i) < 1 - \phi_0$.

## VI. Evaluation Results

We simulate the running example in Section II with a set of $N$=40 service providers to investigate the emergence of semantic interoperability. 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 registered service providers. In our setting, the agents use AMC [7] and COMA++ [8] as two different mapping tools to generate the mappings between the schemas of the 2 registries, and the mappings are given as XML path correspondences, e.g. *(schema1.path1, schema2.path2)*. The best qualified overall schema mapping is defined as a perfect set of mappings, which is provided by an expert. Initially, in each experiment, individual agents autonomously generate an instance of the schema mapping by using the aforementioned mapping tools. While 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 explained below. We perform each experiment in a number of rounds. In each round, each agent decides on 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 $w$; $w$ can be specified in terms of a number of rounds or a number of received advertisements. Each agent evaluates the quality improvement by the Precision [13] of its schema mapping instance, i.e. the proportion of correct mappings in the schema mapping instance of an agent relatively to the mappings provided by the expert. For each experiment, the values of the parameters in the homogeneous setting are considered as the mean values of respective parameters in the heterogeneous setting.

First, we investigate the decision process of adoption in the set of agents. As the ratio of $\gamma$ is the determining parameter for the adoption decision of the agents, we keep the advertisement cost $c_{adv}$ constant and run the experiment for different values

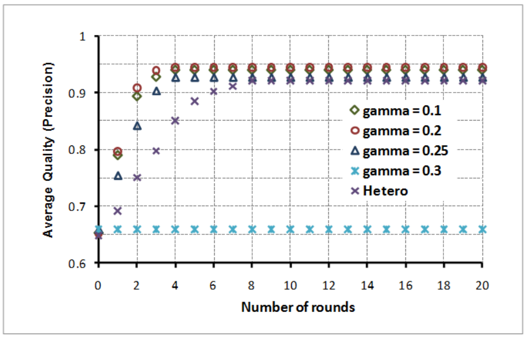


Fig. 1. The effect of increasing $\gamma$ for $c_{adv}$= 0.1$c_a$, $A$=40, $N$=40, $\phi_0$=0.9, and in the heterogeneous setting, $\bar{\gamma}$=0.25

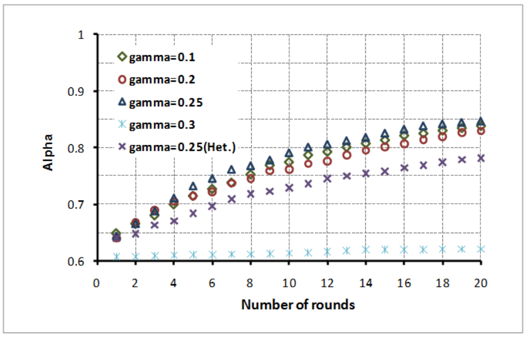


Fig. 2. The evolution of the expected value of $\alpha$ for $c_{adv}$= 0.1$c_a$, $A$=40, $N$=40, $\phi_0$=0.9, and in the heterogeneous setting, $\bar{\gamma}$=0.25

of $\gamma$. The average quality per round is illustrated in Fig. 1 to investigate the convergence rate to emergence of semantic interoperability. As expected, if $\gamma$ 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 and hence, the improvement in the average quality of the schema mapping is insignificant (e.g. see Fig.1 for $\gamma = 0.3$). As experimentally found, the $\gamma$ threshold is between 0.25 and 0.3, which is close to our analytical prediction. Also, the minimum quality $\alpha$ of the schema mapping is approximately around 0.65, as depicted in Fig. 2. Thus, according to our analysis, and specifically Eq. (8), only a value of $\gamma$ lower than 0.286 would result to the convergence of the schema quality to $\phi_0 = 0.9$ or equivalently to the emergence of semantic interoperability. In Fig.2, it is clearly illustrated the sequential increments in the expected value of $\alpha$ satisfy the stated condition in Eq. (10) for low-enough values of $\gamma$. Therefore, the sufficient condition for reaching the Pareto-efficient equilibrium is satisfied for $\gamma$ values lower than 0.286, as discussed in SectionV .

Next, we investigate the decision process of advertisement for the individual agents. We keep the benefit, the adoption cost and the heterogeneity cost constant, while we vary the advertisement cost in different runs of the experiment. As experimentally shown (see Fig. 3), the higher the advertisement cost, the longer the time to reach the same average quality of the schema mapping, since a smaller number of agents decide to take the risk of advertising. It is clear, although not illustrated, that for a relatively high cost of advertisement, there would be no consensus on the quality of the schema mapping, since the agents would avoid to advertise their partial mappings altogether, as we expected by Eq. (5).

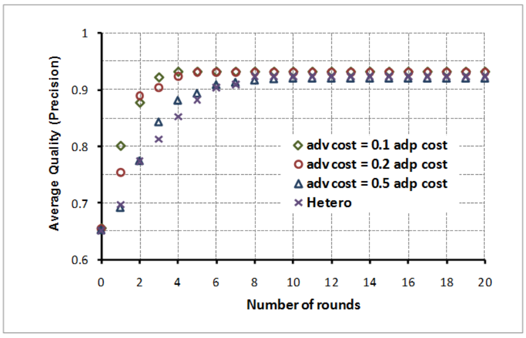


Fig. 3. The effect of increasing advertisement cost for $b$=1, $c_a$=$c_h$=0.1$b$, $A$=40, $N$=40, $\phi_0$=0.9, and in the heterogeneous setting, $\bar{c}_{adv}$=0.1$\bar{c}_a$

## VII. 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. [2] has formalized the notion of interoperability by considering the choices of each autonomous agent and studies how their choices are coordinated. The authors in [14] propose an approach in which the agents use machine-learning techniques to learn new concepts from other agents through instance examples to improve their communication and thus cooperation abilities. Also, the authors in [15] propose to set up semiotic dynamics that achieve distributed coordination. However, in our case, not all the agents necessarily converge to the same semantic mappings, but they can still reach an agreement by establishing a consensus on the quality of the schema mapping. In [16], the authors analyze a dynamic model known as the Naming game and discuss that it is able to account for distributed coordination of autonomous agents and emergence of global agreement. The authors in [17] study semiotic dynamics of a similar model to [16], aiming at defining the microscopic behavior of the agents in emergence of shared vocabularies. They show that the model exhibits the same phenomena as observed in human semiotic dynamics, namely a period of preparation followed by a rather sharp transition into global coordination. However, these developed models cannot be applied in the context of establishing consensus on the quality of a schema mapping, since they involve two-way pairwise interactions, different mapping verification method, and different end of game, as opposed to our model. Moreover, the approach in [18] in the context of collaborative data sharing has addressed semantic interoperability in a very similar setting to the sample scenario of our work. However, the proposed decision making approach in [18] is pertained to *static individual-based* policies for the participants (to adopt or reject updates), while we develop a *dynamic consensus-based* decision making process for the participating agents. In summary, while most of the aforementioned approaches focus on developing different techniques in order to establish interoperability, we focus on investigating under which cost and benefit conditions semantic interoperability emerges, when agents are self-interested.

## VIII. Conclusion

In this paper, we employed a game-theoretic approach to analyze the dynamics of bottom-up 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. As a future work, we plan to investigate the emergence of semantic agreement in case of different agent communication models.

## References


[1] J. J. Jung, "Service chain-based business alliance formation in service-oriented architecture," *Expert Systems with Applications*, vol. In Press, Corrected Proof, 2010.

[2] 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 *Proc. of AAMAS '09*.

[3] G. Vetere and M. Lenzerini, "Models for semantic interoperability in service-oriented architectures," *IBM Systems Journal*, vol. 44, no. 4, pp. 887–903, 2005.

[4] K. Aberer, P. Cudr-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 *Proc. of DASFAA'04*.

[5] D. G. I. for Standardization, *The economic benefits of standardization*. Beuth Verlag, 2008.

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

[7] E. Peukert, J. Eberius, and E. Rahm:, "Amc : A framework for modelling and comparing matching systems as matching processes," in *Proc. of ICDE'11*.

[8] D. Aumueller, H. Do, S. Massmann, and E. Rahm, "Schema and ontology matching with coma++," in *Proc. of the ACM SIGMOD'05*.

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

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

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

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

[13] F. Duchateau, Z. Bellahsne, and E. Hunt, "Xbenchmatch: a benchmark for xml schema matching tools," in *Proc. of VLDB'07*.

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

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

[16] 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*, vol. 19, no. 5, pp. 785–812, 2008.

[17] 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*, 2006.

[18] N. Taylor and Z. Ives, "Reconciling while tolerating disagreement in collaborative data sharing," in *Proc. of ACM SIGMOD'06*.