

# An Empirical Study to validate the Use of Ontological Guidelines in the Creation of $i^*$ Models

Ramilton Costa Gomes Júnior<sup>1</sup>, Renata Silva Souza Guizzardi<sup>1</sup>,  
Xavier Franch<sup>2</sup>, Giancarlo Guizzardi<sup>1</sup>, Roel Wieringa<sup>3</sup>

<sup>1</sup>Ontology and Conceptual Modeling Research Group (NEMO).  
Federal University of Espírito Santo, Vitória/ES, Brasil.

<sup>2</sup>Universitat Politècnica de Catalunya (UPC)  
Barcelona, Spain

<sup>3</sup>University of Twente  
Enschede, The Netherlands

*ramiltoncosta@gmail.com, {rguizzardi, gguizzardi}@inf.ufes.br,  
franch@essi.upc.edu, roelw@cs.utwente.nl*

**Abstract.**  $i^*$  is a well known goal modeling framework, developed by a large and geographically dispersed research community. Currently,  $i^*$  users tend to ascribe different and conflicting meanings to its constructs, leading to a non-uniform use of the language, and consequently undermining its adoption. In previous works, we proposed ontological guidelines to support the creation of  $i^*$  models, in an attempt to provide a solution to this problem. In this paper, we present an empirical study, to evaluate these ontological guidelines. Results show that for more experienced conceptual modelers, the ontological guidelines indeed support  $i^*$  modeling. However, results are not as positive for non-experienced conceptual modelers.

## 1. Introduction

$i^*$  is a goal modeling framework used for Requirements Engineering (Yu, 1995). In the past twenty years, this framework has attracted the attention of different research groups, which have-proposed different variants of the initial framework, each one proposing different semantics to the language's constructs. The community that develops  $i^*$  is aware that this non-uniform use of  $i^*$  makes it difficult for novices to learn how to use the language, besides undermining its acceptance in industry.

We believe this problem can be solved with the use of a foundational ontology to interpret the semantics of the  $i^*$  concepts. A foundational ontology is a formal system of domain-independent categories that can be used to characterize the most general aspects of concepts and entities that belong to different domains in reality (Guizzardi, 2005). The idea is to apply the foundational ontology as a reference model to interpret the concepts of the language. Then, based on such interpretation, we are able to provide some guidelines to support modeling, here referred as *ontological guidelines*. In previous works (Guizzardi, Franch, Guizzardi, 2012), (Guizzardi, Franch, Guizzardi, Wieringa, 2013), we proposed some ontological guidelines for  $i^*$  modeling, based on

the UFO foundational ontology (Guizzardi, 2005) (Guizzardi et al, 2013)(Guizzardi, Falbo, Guizzardi, 2008). The aim of this paper is to present the experimental design and the results of an empirical study conducted to evaluate the use of such ontological guidelines.

Nowadays, empirical studies are considered appropriate means to prove the effectiveness of a new approach. For (Vokac, 2002), the ideal science would have a set of empirical observations for each theory, either to support the theory or to prove it wrong. In other words, empirical observation is the core of the scientific process. Furthermore, it is through empirical observation that one can check theories, explore critical factors and give light to new phenomena, so that the theories can evolve (Travassos, 2002).

Having this in mind, we decided to conduct an experiment to confirm our intuitions that the use of ontological guidelines lead to  $i^*$  models of better quality. The experiment was conducted in two colleges, having fifty-five subjects in total. The subjects were students of a Systems Analysis and Development course and the PhD and Master program in Computer Science. The main goal of the experiment was to verify if the ontological guidelines cited above are useful or not in the development of  $i^*$  models. For that, the subjects participated in modeling activities with and without the use of the guidelines and, then, the results were compared. In the experiment applied with PhD and master students, the results show that the ontological guidelines are useful for the development of  $i^*$  models. Among the population of the second experiment application, composed of less experienced conceptual modelers, the experiment results were not so positive.

The remainder of this article is organized as follows: Section 2 presents information on the  $i^*$  framework and its variants; Section 3 describes the UFO fragment applied in this work; Section 4 presents some of the proposed ontological guidelines; Section 5 describes the empirical study; and, finally, Section 6 concludes the paper.

## **2. The $i^*$ Framework and its Variants**

The original  $i^*$  framework is described in (Yu, 1995). Since then, several variants have been proposed, for instance GRL and Tropos, see (Cares, 2012) for an overview. Some variants come from paradigm shifts, others propose some particular type of new construct, and still others issue slight modifications related to the core constructs of the  $i^*$  language.

One of the most controversial constructs in the language is the means-end link. In the original  $i^*$  (Yu, 1995), this link is used to connect a goal or a task to softgoals. In GRL, this link is applied to connect a task to a goal, a task to a task and a resource to a task. However, in the  $i^*$  wiki, one of the major sources of material about the language, this link is only used to connect a task to a goal. (Cares, 2012) also points out that different versions of Tropos propose different uses for the means-end link.

These different interpretations and uses make a new  $i^*$  learner confuse. She may ask herself: when can I use a means-end link after all? Why is it used this way? Why can't I use a means-end link between a resource and a goal, for example? We argue that

the best way to respond to these questions is to understand the ontological semantics behind the constructs of the language. By understanding their ontological nature, we may provide good reasons why a concept or a link may or may not be used in a particular way.

### **3. Background: The UFO Foundational Ontology**

Here we briefly present the UFO concepts that are used in this paper provide an interpretation to  $i^*$ . To facilitate reading we use a different font to highlight the UFO concepts. For a fuller presentation on UFO, the reader should refer to (Guizzardi, 2005), (Guizzardi et al, 2013) and (Guizzardi, Falbo, Guizzardi, 2008).

In UFO, a stakeholder is represented by the `Agent` concept, defined as a concrete `Endurant` (i.e. an entity that endures in time while maintaining its identity) which can bear certain `Intentional States`. These intentional states include `Beliefs`, `Desires` and `Intentions`. `Intentions` are mental states of `Agents` which refer to (are about) certain `Situations` in reality. `Situation` are snapshots of reality. The propositional-content (i.e., `proposition`) of an `Intention` is termed a `Goal`.

In contrast to `Endurants`, `Events` are perduring entities, i.e., entities that occur in time accumulating their temporal parts. `Events` are triggered-by certain `Situations` in reality (termed their pre-situations) and they change the world by producing a different post-situation. `Action` are deliberate `Events`, i.e., `Events` deliberately performed by `Agents` in order to fulfill their `Intentions`. An `Action` achieves a `Goal` if the `Action` brings about a `Situation` in the world that satisfies that `Goal`.

In contrast with an `Agent`, an `Object` is a concrete `Endurant` that does not bear intentional states or perform actions. An `Object` participating in an `Action` is termed a `Resource`.

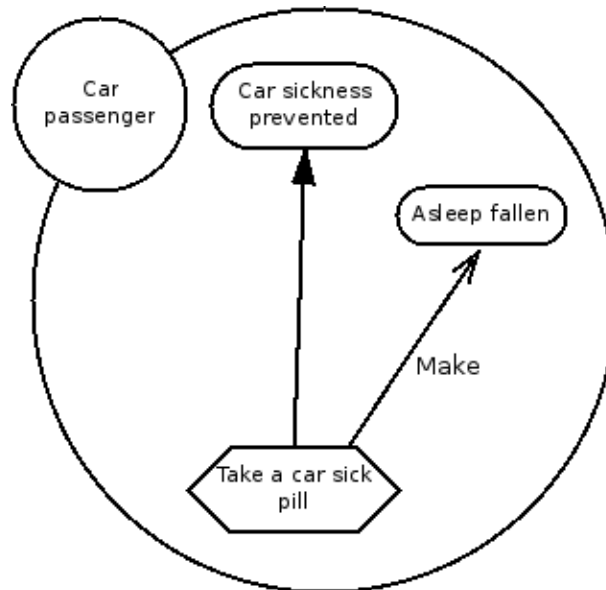
### **4. Ontological Guidelines for the Creation of $i^*$ Models.**

In this section, we describe some of the proposed ontological guidelines. For lack of space, we are not able to present them all and refer to (Guizzardi, Franch, Guizzardi, 2012) and (Guizzardi, Franch, Guizzardi, Wieringa, 2013) for a full description. In total, there are seven ontological guidelines and all of them have been considered in the experiment.

First, it is important to point out that we interpret  $i^*$  goals, tasks, resources and agents as their counterparts in UFO (with `Action` as task). Having that in mind, let us try to interpret the  $i^*$  decomposition relation. Since goals are propositions, due to its ontological nature, it is impossible for a goal to be decomposed into tasks or resources. Thus, goals can only be decomposed into subgoals. Consequently, when decomposing goals, an  $i^*$  *and-decomposition* is interpreted as a conjunction of subgoals, while an  $i^*$  *or-decomposition* is interpreted as a disjunction of subgoals. Similarly, softgoals, tasks and resources can only be decomposed into softgoal, tasks and resources, respectively. This originates the ontological guideline describe in the first line of Table 1.

In  $i^*$ , a means-end link is applied to connect a means to an end. For example, a task T (means) to a goal G (end), meaning that the execution of T leads to the achievement of G. Here, we adopt the conceptual modeling evaluation method proposed in (Guizzardi, 2005) that states that we should avoid construct redundancy, i.e., two language constructs should not be applied to model the same phenomenon in the world. Construct redundancy adds unnecessary complexity to the modeling language, besides making specifications more difficult to understand. Moreover, when facing redundancy, designers tend to ascribe slightly different meanings to the redundant constructs, which may not be fully understood by the model readers. In our case, if we allow, for instance, goals G2 and G3 to be connected via means-end to goal G1, we will not be able to differentiate between means-end and or-decomposition, i.e. these two links will be applied to represent the very same relation in the world. Thus, this will be a case of construct redundancy. To avoid that, we propose the ontological guideline described in the second line of Table 1.

In  $i^*$ , a make-contribution is applied between a task T and a goal G, meaning that if T is executed, then G is fully achieved. But if this is so, how can one differentiate between means-end and make-contribution? Using UFO, we differentiate this by looking at the intention behind the execution of T. To understand this, let us consider the  $i^*$  model depicted in Figure 1, which exemplifies the use of the means-end and the make contribution links.



**Figure 1. Means-end vs. make-contribution**

In Figure 1, a Car Passenger<sup>1</sup> agent executes the Take a car sick pill task in order to prevent himself from being sick during the journey he is making (means-end link to Car sickness prevented goal). As a side effect of this medication, the Car Passenger also goes to sleep (make-contribution link to Asleep fallen goal).

As result of the mapping from  $i^*$  tasks into UFO actions, every task is associated with a motivating intention whose propositional content is a goal. In other words, we

<sup>1</sup> From now on, we use a different font for the names of the instances of the  $i^*$  actors and intentional concepts, such as goals, tasks, and resources.

execute a particular task in order to accomplish a specific goal. In  $i^*$ , the association between the task and the goal in this case is made by a means-end link (e.g. **Take a car sick pill** task as means to **Car sickness prevented** goal). On the other hand, this same task can also generate some other goals to be accomplished, without however, being intended by the choice of this particular task. In this case, a make-contribution link is established (e.g. **Take a car sick pill** task contributing to **asleep fallen** goal). In other words, the means-end link or the make-contribution link should be applied according to the ontological guideline described in the third line of Table 1.

**Table 1. Some of the proposed  $i^*$  ontological guidelines**

<b><i>Ontological Guidelines</i></b>
1. A decomposition link can only be applied between elements of <i>the same kind</i> . E.g. goal->goal, task->task.
2. A means-end link can only be applied between elements of <i>different kinds</i> . E.g. task->goal, resource->task.
3. Taking task T and goal G, if the intention behind the execution of task T is to accomplish G, T and G should be related via <i>means-end link</i> . On the other hand, if by executing T, G is unintentionally achieved (i.e., as a side-effect of the execution of T), then T and G should be related via <i>make-contribution</i> .

## **5. The Empirical Study**

In this section, we describe the empirical study we conducted to evaluate the use of the ontological guidelines. The hypothesis of the study is "*the ontological guidelines enhance the capability of the subjects to create  $i^*$  models.*" The experiment was conducted in a controlled environment and is based on a quantitative strategy, in which the data is analyzed using statistical and descriptive methods. For the experimental design, we followed the framework presented in (Kochanski, 2009).

### **5.1 Experimental Design**

The experiment has as object of study two  $i^*$  models (here referred to as Case 1 and Case 2), representing two different situations. Each participant had to complete the models, by filling in the blanks with the correct element or link to be used in each question. Figure 2 illustrates part of one model. For each blank, there are two and more possibilities, having as alternatives constructs of  $i^*$  whose use normally generates confusion or doubts. For example, in Question 2 (refer to Figure 2), the participants should indicate if "Provide gift wrapping solution" is a goal or a plan. In Question 5, the participants should indicate if "Provide gift wrapping solution" and the two tasks "Organize wrapping stand" and "Allow vendors to wrap gifts" should be linked via OR-means-end or via OR-decomposition. The idea is to verify if the participants can select them intuitively (pre-test) or if the use of ontological guidelines (post-test) effectively helps the selection of the correct construct.

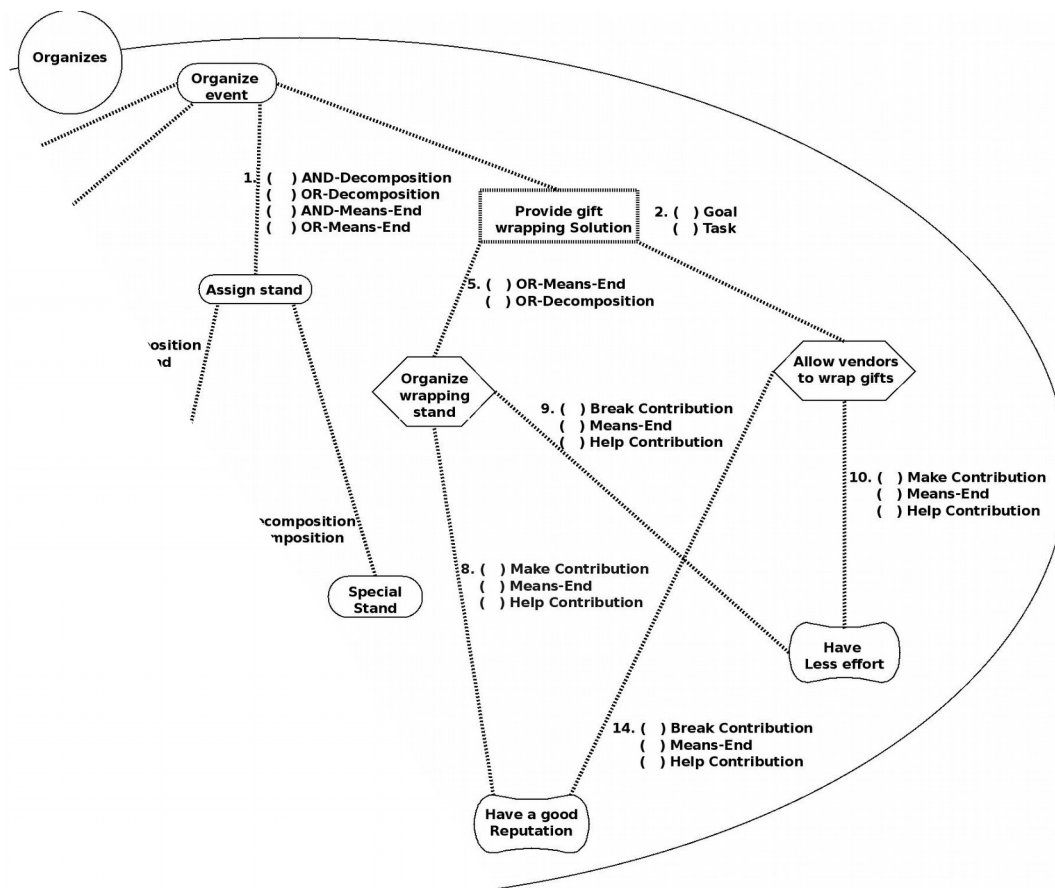


Figure 2. Part of the  $i^*$  model

The experiment was divided in two steps: pre-test and post-test. In the pre-test, all participants performed the first activity, i.e. filling in the blanks, using Case 1. Then, in a separate form, they justified their choices for each blank. During this activity, all participants had a printout of some slides containing basic information about  $i^*$  (the  $i^*$  wiki guidelines), as well as the description of Case 1. No information about the guidelines is given in this first step.

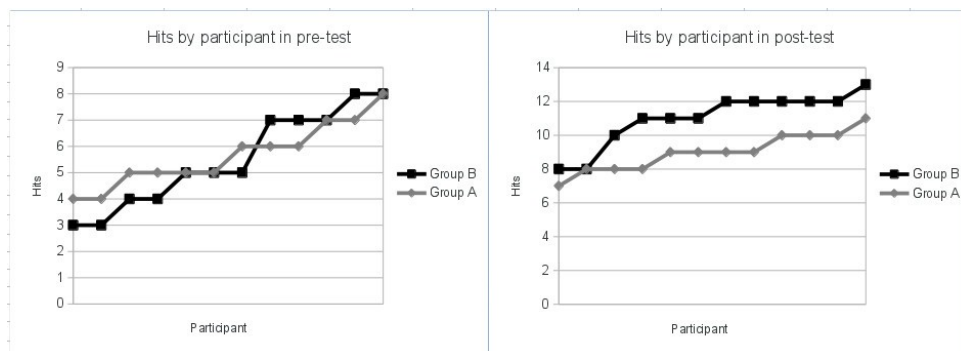
After the pre-test activity, the students were randomly divided into two groups: group A (control group) and group B (experimental group). After the division, the participants of group A moved to another room to perform the post-test activity. Both groups had to perform a second activity of filling in the blanks, now using Case 2. However, in this part, only group B received information about the ontological guidelines. Both groups had the description of Case 2 and group B also had a printout of some slides containing the ontological guidelines. In the post-test, the participants of both groups were also asked to fill in a separate form justifying their choices for each blank.

To capture the impression of the participants about the guidelines, the participants were also asked to respond some questions regarding their opinion about the  $i^*$  wiki guidelines and the ontological guidelines.

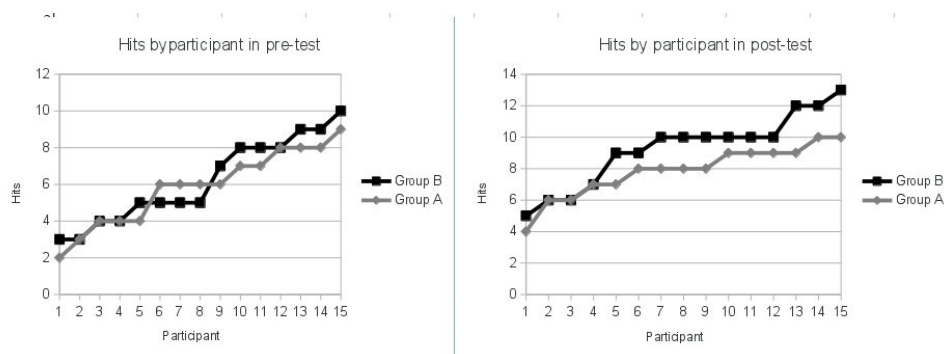
## 5.2 Collected Data

The data was collected through questionnaires. Before the experiment activities, we applied a questionnaire to capture the participants' profile. We applied the experiment twice, with two different populations. We will here refer to these applications as application 1 and application 2. In application 1, there were 24 participants: 16 of them were undergraduate students of Computer Science or Computer Engineering, 7 of them were master students in Computer Science, and 1 of them was a PhD student in Computer Science. The participants were assigned into two groups of 12 participants, which were balanced in terms of educational level and modeling experience. In both groups, there was one participant with 1-3 years of experience in goal modeling and  $i^*$ , while the others declared not having experience in this area. In application 2, there were 30 participants, all of them in the final year of an undergraduate course in Information Systems Analysis and Development. Each group had 15 participants. None of the participants indicated having experience in goal modeling or  $i^*$ .

Both in the pre-test and in the post-test, the same activities and questionnaires were used in applications 1 and 2. The graphs of Figures 3 and 4 show the results for the first and the second application, respectively. When the participant fills in the blank correctly, we say that he has a hit.



**Figure 3. Hits by participant in pre-test (left) and post-test (right) in the first experiment application.**



**Figure 4. Hits by participant in pre-test (left) and post-test (right) the second experiment application.**

Tables 2 and 3 present data regarding the number of hits per participant in the first and second application, respectively. The columns present data on average, median, highest and lowest value of number hits per participants.

**Table 2 - Number of hits per participants in the first application**

	Average		Median		Highest		Lowest	
	Group A	Group B	Group A	Group B	Group A	Group B	Group A	Group B
Pre-test	6,67	5,50	5,50	5,00	8,00	8,00	4,00	3,00
Post-test	9,00	11,00	9,00	11,50	11,00	13,00	7,00	8,00

**Table 3 - Number of hits per participants in the second application**

	Average		Median		Highest		Lowest	
	Group A	Group B	Group A	Group B	Group A	Group B	Group A	Group B
Pre-test	5,87	6,20	6,00	6,00	5,00	10,00	2,00	3,00
Post-test	7,89	9,27	9,27	8,00	10,00	13,00	4,00	5,00

### 5.3 Data Analysis

Analyzing Figure 3, we notice that in the pre-test of the first application, the participants of group A scored a larger number of hits than the participants of group B. However, in the post-test, group B performed better than group A. This shows that the group that used the ontological guidelines performed better when compared to the group that only had access to the *i\** wiki guidelines. This result favors our hypothesis, supporting the idea that the ontological guidelines effectively help the creation of *i\** models.

By looking at Figure 4, we see that in the pre-test of the second application, groups A and B showed a great balance in realizing the activities; both groups scored the same number of hits and errors. In the post-test, group B achieved a significantly higher number of hits in relation to group A, as seen in Figure 4. Again, this result favors our hypothesis, supporting the idea that the ontological guidelines effectively help the creation of *i\** models.

Table 2 shows the data regarding the number of hits per participants in the pre-test and post-test, in the first application. The values for average, median, highest and lowest are very similar in the pre-test activity. But in the post-test activity, the values are significantly different, result that favors ours hypothesis.

Table 3 presents the data regarding number of hits per participants in the pre-test and post-test, in the second application. The values for average, median, highest and lowest have a small difference in the pre-test activity. But in the post-test activity, the values are significantly different, result that favors ours hypothesis.

The descriptive analysis we presented so far is able to provide us with some evidence supporting the hypothesis, We can quantify this support by a statistical test. Thus, we also applied the Wilcoxon-Mann-Whitney statistical test, with a significance level of 5%, to compare the hits for each participant between the experimental (group



B) and control (group A) groups, in both experiment applications. This statistical method is a non-parametric method recommended for small samples or groups with less than 20 participants (Robson 2002). In the first application, the calculated U value is 23 and the critical U value from the Mann-Whitney index is 37. Since the calculated U is lower than the critical U, then we may conclude that the values are significantly different between the groups, which supports our hypothesis. In the second application, the calculated U value is 65 and the critical U value from the Mann-Whitney index is 64. Since in this case, the calculated U is not lower than the critical U, then we cannot confirm our hypothesis.

Given the results of the Mann-Whitney test, we cannot conclude that the ontological guidelines are always helpful. We attribute this difference to the divergence in profiles in the two experiment applications. The participants of the first application have a higher graduation level than the participants of the second application, and thus are, in general, more experienced in conceptual modeling. Thus, we claim that the ontological guidelines are helpful for more mature conceptual modelers. New empirical studies should be conducted to confirm this hypothesis.

Regarding the qualitative evaluation of the ontological guidelines, we have the following results. In the first application, 7 out of 12 participants considered that the ontological guidelines are better than the *i\** wiki guidelines. The other 5 participants considered that the ontological guidelines and the *i\** wiki guidelines have the same quality. When asked about the usefulness of the ontological guidelines, 8 participants considered them very useful, 2 participants found them not very useful and 2 participants found them indifferent. In the second application, 13 out of 15 participants considered that the ontological guidelines are better than the *i\** wiki guidelines, while 2 participants considered that the ontological guidelines and the *i\** wiki guidelines have the same quality. Regarding the usefulness of the ontological guidelines, 10 participants found them very useful, 3 participants found them not so useful and 2 found them indifferent. We find these results positive, as most of the participants had a good perception regarding the ontological guidelines.

Let us now analyze which questions were more difficult, i.e. led to more errors in both experiment applications. This will allow us to find out which ontological guidelines are not clear and should be improved. In the first application, the questions that led to more errors were questions 8 and 10. In the second application, the questions that led to more errors were questions 7, 9 and 14. Questions 8, 9, 10 and 14 regard the use of the means-end, make-contribution and help-contribution links. We conclude that the participants in both experiment applications could not understand well the ontological difference between these three links. Thus, the ontological guidelines concerning this differentiation should be improved. Question 7 regards the differentiation among AND and OR decomposition. We conclude that in the second application, the participants also had doubts regarding the use of decomposition. Thus, the guidelines concerning these links should also be improved.

#### **5.4 Threats to Validity**

The following factors are considered the main threats to the validity of this empirical

study:

- a) the heterogeneity of the participants of the first application, since they had different academic degrees. To mitigate this risk, we collected information about the academic degree of the participants in the profile questionnaire and took this into account in our experiment design;
- b) the possibility that the participants had previous knowledge of the ontological guidelines. To remediate this risk, we asked in the experiment questionnaire if the participant had had previous contact with the guidelines. This information was taken into account in our analysis;
- c) the chance that the participants had low interest in the experiment results, carelessly performing the experiment activities. To mitigate this risk, we tried to motivate the participants, showing the importance of the results of the experiment. Moreover, the experiment was designed to be as short as possible, so as to prevent tiredness and disinterest;
- d) the possibility that the researcher conducting the experiment influenced the experiment results. To remediate this risk, the researcher conducting the experiment tried to be as objective and unbiased as possible during the experiment activities;
- e) the possibility that the subjects had a positive opinion about the guidelines, because they knew we were the ones who formulated them. To remediate this risk, we did not tell them we were the authors of the guidelines.

## 6. Final Considerations

This article presented an empirical study with the objective to evaluate the use of ontological guidelines to create  $i^*$  models. For that, the experiment was conducted in two steps (pre and post-test), in which the participants performed modeling activities without (pre-test) and with (post-test) the use of ontological guidelines. To analyze the results, we performed the Mann-Whitney statistical test. The outcome supports our hypothesis that states that the guidelines are useful, and does not provide evidence against it. Moreover, most participants stated that they found the ontological guidelines useful to support them in the creation of  $i^*$  models.

Given the results of this experiment, we intend to develop an  $i^*$  modeling tool that uses the ontological guidelines as support for the model designer. For that, we aim at proposing a metamodel that is compatible with these guidelines, to serve as basis for the development of the tool.

For the future, we also intend to perform new experiments to collect more data regarding the use of the ontological guidelines to create  $i^*$  models. In order to confirm our hypothesis, we must repeat the designed experiment, taking populations of different profiles. We aim, for example, to conduct the experiment with professional modelers. Moreover, we intend to perform different experiments. For instance, we would like to conduct an experiment in which the participants are asked to create  $i^*$  models from scratch, with and without the use of the ontological guidelines. Then, based on some pre-established criteria collected from  $i^*$  experts, we will be able to analyze if the

models created with the use of ontological guidelines have higher quality than the ones created without them.

**Acknowledgement.** This work is partially supported by CAPES/CNPq (grant number 402991/2012-5), CNPq (grant numbers 461777/2014-2 and 485368/2013-7), and the Spanish project EOSSAC, ref. TIN2013-44641-P.

## References

- Ayala, C., Cares, C., Carvalho, J.P., Grau, G., Haya, M., Salazar, G., Franch, X., Mayol, E. and Quer, C. (2005), “A Comparative Analysis of  $i^*$ -Based Agent-Oriented Modeling Languages”, In: 17th International Conference on Software Engineering and Knowledge Engineering, Taipei, Taiwan, pp. 43-50.
- Cares, C. (2012), “From the  $i^*$  Diversity to a common interoperability framework”, PhD Thesis, Software Engineering for Information System Research Group, UPC, Spain.
- Guizzardi, G. (2005), “Ontological Foundations for Structural Conceptual Models”. PhD Thesis, University of Twente, The Netherlands.
- Guizzardi, G., Wagner, G., Falbo, R.A., Guizzardi, R.S.S., Almeida, J.P.A. (2013), Towards Ontological Foundations for the Conceptual Modeling of Events, 32nd International Conference on Conceptual Modeling (ER 2013), Hong Kong.
- Guizzardi, G., Falbo, R. A., Guizzardi, R. S. S. (2008), Grounding Software Domain Ontologies in the Unified Foundational Ontology (UFO): The case of the ODE Software Process Ontology , 11th Iberoamerican Conference of Software Engineering (CIbSE 2008), Recife, 2008.
- Guizzardi, R., Franch, X. and Guizzardi, G. (2012), “Applying a Foundational Ontology to Analyze Means-end Links in the  $i^*$  Framework”, In: 6<sup>th</sup> IEEE International Conference on Research Challenges in Information Science, Valencia Spain, pp. 1-11.
- Guizzardi, R., Franch, X., Guizzardi, G. and Wieringa, R. (2013), “Ontological Distinctions between Means-end and Contribution Links in the  $i^*$  Framework”, Lecture Notes in Computer Science, v. 8217, Heidelberg: Springer, pp. 463-470.
- Kochanski, D. (2009), “Um Framework para Apoiar a Construção de Experimentos na Avaliação Empírica de Jogos Educacionais”, Master Dissertation in Applied Computing, UNIVALI, Brazil.
- Lucena, R., Santos, B., ; Silva, J., Silva, L., Alencar, R. and Castro, B. (2008), “Towards a Unified Metamodel for  $i^*$ ”. In: 2nd IEEE International Conference on Research Challenges in Information Science, Marrakech, v. 1, pp. 237-246.
- Santos, B. (2008), “Istar Tool – Uma proposta de ferramentas para Modelagem de  $i^*$ ”, Master Dissertation in Computer Science, UFPE, Brazil.
- Travassos, G. (2002), “Relatório Técnico RT-ES-590/02 – Introdução à Engenharia de Software Experimental”. Systems Engineering and Computer Science Program. COPPE/UFRJ, Brazil.

- Vokac, M. (2002), "Empiricism in Software Engineering: A Lost Cause?" Essay for MNVIT401.
- Yu, E. (1995), "Modelling Strategic Relationships for Business Process Reengineering", Ph.D. thesis, Dept. of Computer Science, University of Toronto, Canada.
- Yu, E. (1997), "Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering", In: 3rd IEEE International Symposium on Requirements Engineering. Annapolis, USA, pp. 226-235.