2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement

Assessing the Intuitiveness of Qualitative Contribution Relationships in Goal Models: an Exploratory Experiment

Sotirios Liaskos
School of Information Technology
York University
Toronto, ON, Canada
liaskos@yorku.ca

Alexis Ronse
School of Information Technology
York University
Toronto, ON, Canada
aronse@yorku.ca

Mehrnaz Zhian
Department of Computer Science
York University
Toronto, ON, Canada
mehrnaz@cse.yorku.ca

***Abstract*—[Background]: Developing conceptual models is an integral part of the requirements engineering (RE) process. Goal models are requirements engineering conceptual models that allow diagrammatic representation of stakeholder intentions and how they affect each other. A specific goal modeling language construct, the contribution of goal satisfaction of one goal to another, plays a central role in supporting decision problem exploration within goal models. [Aims]: We report on an experiment whose aim was to measure the user perception of the meaning of the aforementioned modeling construct. [Method]: A set of contributions under different scenarios were given to experimental participants who were asked what they thought the effect of the contribution was. [Results]: We found that participants are not always in agreement either within themselves or with the designers' intentions on the meaning of the language. [Conclusions]: The results call for possible adaptations to the way goal modeling languages are used.**

***Index Terms*—Conceptual modeling, requirements engineering, goal models, model comprehension**

1. Introduction

Developing conceptual models lies at the heart of the Requirements Engineering (RE) process. By building conceptual models, analysts are able to organize problem information in a way that facilitates a variety of activities such as validation, analysis and design. Conceptual models in RE are typically visualized using box-and-line diagrammatic notations which are aimed at efficiently capturing and communicating large amounts of information.

Goal models [1], [2] are diagrammatic requirements engineering conceptual models that have been found to be useful for capturing the structure of stakeholder intentions during the requirements analysis process. At their core, goal models contain intentional elements (i.e., goals of various kinds) and relationships between them (e.g., refinement or satisfaction influence), diagrammatically visualized as boxes and lines, respectively. Several goal modeling languages have been proposed since the introduction of the concept, including *i** [1], KAOS [2], [3], Tropos [4] and URN/GRL [5]. One of the key constructs within many of those goal modeling languages is the satisfaction *contribution* relationship, which represents how the satisfaction of one goal affects the satisfaction (or non-satisfaction, i.e., denial) of another goal. Several ways for modeling, interpreting and visualizing the corresponding construct have been proposed in the RE literature [6], [7], [8], [9], [10]. Criteria such as expressiveness or well-formedness are traditionally used to evaluate the effectiveness of a proposed way to represent and assign semantics to the construct. However, there is limited work on how the intended meaning of the construct aligns with how users perceive and use it.

In this paper, we focus on a particular way to define the appearance and meaning of contributions in goal models, and experimentally evaluate it using human participants. Experimental participants are exposed to various contribution links and scenarios of satisfaction levels of the origin goal and are asked to identify the satisfaction of the destination goal. The results indicate, among other things, that negative contribution and satisfaction levels appear to cause more disagreement and deviation from the normative semantics.

The rest of the paper is organized as follows. In Section 2 we present goal models and contribution links. In Section 3 we describe our experimental design and results. We explore related work in Section 4 and conclude in Section 5.

2. Background

2.1. Goal Models and Contribution Links

The goal models that we focus on in this paper look like the one of Figure 1 and they are akin to models of the *i** goal modeling language ([1] for a baseline). The boxes in the diagram represent two types of intentional elements. The ovals represent *hard-goals*, i.e. states of affairs that an actor in question wants to bring about, avoid or maintain. The cloud-shaped elements are *soft-goals*. Soft-goals differ from hard-goals in that they do not have a clear-cut criterion for deciding if they are satisfied or not. Thus, in goal models we assume that soft-goals are satisfied to a certain degree based on evidence external to the model or based on our knowledge of satisfaction of other goals in the model.

978-1-5090-4039-1/17 $31.00 © 2017 IEEE
DOI 10.1109/ESEM.2017.69

IEEE computer society

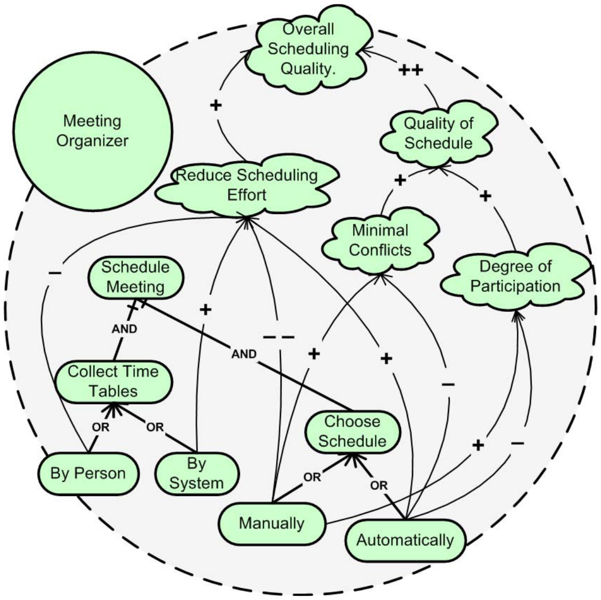

Figure 1. A Diagrammatic Goal Model

To show how our knowledge of satisfaction of a goal affects satisfaction of another goal we use contribution links. Contribution links can be positive or negative and can have various degrees of intensity. Several ways for modeling contribution links have been proposed, visualized through appropriately labeling the link. Numeric approaches use numbers as labels [6], [9], [8], [10]. In such approaches, a larger number represents a more intense contribution. A sign may also be added to show if the contribution is positive or negative [6], [9]. Another approach, which is featured in the original *i** [1] and GRL [6] goal modeling languages, are qualitative contribution links in which labels are presented through one of the symbols “++”,“+”,“−”,“−−”. Figure 1 features links with such labels. Each of these symbols means that satisfaction of the origin of the contribution *makes*, *helps*, *hurts* or *breaks* satisfaction of the destination, respectively. In this paper, we focus on this qualitative type of contribution link.

2.2. Propagation Semantics

Knowledge of the informal meaning of the above choices of contribution labels can be claimed to be useful for roughly assessing the satisfaction status of various goals in a goal model, for e.g. identifying optimal alternatives. However, more formal approaches for defining the semantics of goal satisfaction have been proposed. Giorgini et al. offer one of the most expressive formalizations of goal satisfaction [11], [9]. Each goal is associated with two variables, a satisfaction and a denial one. Each of the two variables takes one of three values: **N**, **P** and **F**, meaning no, partial and full satisfaction or denial, respectively. For convenience we use the symbols **FD** (full denial), **PD** (partial denial), **N** (no information), **PS** (partial satisfaction), and **FS** (full satisfaction) to denote different levels of each of the variables.

TABLE 1. Contribution Semantics

Label	Effect	Label	Effect
++	FS → FS	−−	FS → FD
	PS → PS		PS → PD
	PD → PD		PD → PS
	FD → FD		FD → FS
+	FS → PS	−	FS → PD
	PS → PS		PS → PD
	PD → PD		PD → PS
	FD → PD		FD → PS

A set of logical rules allows the inference of the satisfaction and denial values of a goal that is pointed by a contribution link given the label of the link and the satisfaction and denial value of the origin. For the interest of space we informally describe the effect of these rules in Table 1, referring the reader to one of [11], [9] for a detailed account. In short, positive labels propagate satisfaction and denial values to satisfaction and denial values, respectively, while negative labels do the reverse. Moreover, while double labels (“++” and“−−”) propagate values as they are, single labels “truncate” full values (**F**) into partial (**P**). Finally, not seen in the table, **N** → **N** for any kind of label.

2.3. Expert Intent versus User Perception

Given its well-formedness, the above formalization is useful for automated reasoning about goal satisfaction – Giorgini et al. indeed introduce a label propagation algorithm for inferring satisfaction values of goals given labels of other goals. However, does the way the formalization works agree with how humans expect to use symbols such as “+”,“++”,“−” and “−−”?

More generally, we use the working term *intuitiveness* to describe the alignment between the meaning that the language designers intend a language feature to have and the meaning that the users infer from the way it is visualized – and possibly from other sources. A way to appreciate this concept is through an analogy with a recognized principle of human-computer interaction design: interaction designs are understood to be more successful when they steer users towards developing a mental model that aligns with the designer's intention of how their artifacts are supposed to work [12]. In conceptual models, both the semantics of modeling constructs and the way they are visualized are available for designers to be chosen in a way that the latter successfully evoke the intended understanding of the former with limited need for training and enforcement. Note here that the designer intent is observed through the design artifact they produce, i.e., we assume the designers have successfully materialized their intent into the actual language design.

To empirically measure intuitiveness, one can test how often the inferences performed by the user – in our case assumptions about satisfaction propagation – agree with inferences prescribed by the designed semantics. In our experiment, we focus on scenarios containing simple contribution links connecting two goals under various assumptions on the satisfaction of the origin goal. We collect the response of participants on what the perceived degree of satisfaction of the destination goal is. Hence we indirectly observe the semantics they assign to the contributions as visualized and roughly explained beforehand.

3. Experimental Study

3.1. Experimental Design

The goals of our study are to explore: (a) if there is agreement among participants as to what the meaning of contribution labels is, (b) if the responses given by the participants agree with the normative formal semantics described earlier, (c) what factors affect the above agreement measures, including label visualization, sign and intensity.

To fulfill the above objectives we developed a set of twenty (20) exercises. Each exercise contains two goals connected with each other using a contribution link. All four (4) kinds of contribution links are considered, initially visualized symbolically: "++","+","−" and"−−". For each kind of contribution link five (5) different scenarios are created. In each of the five scenarios the origin goal is satisfied by each of the five possible degrees discussed above: **FD**, **PD**, **N**, **PS**, **FS**. It is, thus, assumed for simplicity that only one of satisfaction or denial variable has a non-**N** value, i.e., there are no conflicts. The satisfaction degree is displayed as a label next to the corresponding goal. Thus, the 20 produced examples are all 4×5 possible combinations of contribution labels and origin satisfaction values.

Each model is placed in a separate screen of an on-line survey-like instrument. In each screen we ask the participants to look at the model and respond with what they think the satisfaction value of the destination goal is. They choose from an inventory of all 5 possible satisfaction values presented in a consistent order. Below that, the participants are asked to rate their confidence in their responses using a Likert-type scale from *Very Unconfident* to *Very Confident*. The screens are given to them in a random order.

Prior to performing the exercises the participants watch a set of instructional videos introducing them to goal models as devices for making decisions, followed by a series of decision making exercises. In the instructional videos the concept of contribution link is explained informally but without any explicit presentation of concrete semantics. For simplicity, satisfaction status of a goal is presented as a unique variable getting values from **FD** to **FS**. The videos also discuss that double labels ('++', '−−') imply stronger influence and that negative labels ('−', '−−') imply negative influence. Precise rules, however, are not given.

Having created the initial instrument this way we then duplicated it. The duplicate instrument has the exact same information except for the fact that labels "+","++","−","−−", are replaced by the actual words *help*, *make*, *hurt*, *break*. The replacement takes place in all 20 models as well as in the instructional videos. This introduces an additional factor in our design, the representation format, with two levels, *symbolic* and *textual*.

Participants are students of York University, taking a first year undergraduate management course. Their participation is solicited through a participant pool system, and they are offered bonus grade for their participation. The instrument is administered in a computer lab in three sessions of the same day and evening time of three consecutive weeks. Participants are split randomly between the two instruments (symbolic and textual).

3.2. Results

3.2.1. Sample Characteristics. A total of forty-one (41) persons showed up for the experiment. Twenty-one (21) participants are randomly assigned to the symbolic instrument, and twenty (20) to the textual instrument. Out of the 41 persons, five (5) are filtered out due to low performance in questions testing their comprehension of concepts presented in the videos and one (1) due to an incomplete response. The remaining thirty-five (35) cases (18 symbolic and 17 textual) are used for the analysis. They are nine (9) males and twenty-six (26) females, their ages are predominantly 18-29 and their field of study primarily Business and Economics.

3.2.2. Analysis Approach. To perform data analysis we consider one between-subjects factor, which is whether we use symbolic or textual representation and two main within-subjects factors: the sign of the contribution (positive or negative) and the satisfaction status of the origin goal (satisfied or denied). We conduct MANOVA [13], [14] to identify statistically significant main or interaction effects.

The response variables we consider are: (a) the total distance between participant responses – to assess agreement and its affecting factors, (b) total relative and absolute distance of the participant responses and the normative response (as per Table 1) – to assess alignment between visualization and semantics, (c) average participant confidence in their response.

3.2.3. Agreement between Participants. To measure the distance between participant responses we first code **FD**, **PD**, **N**, **PS**, **FS** (the observed responses) to the interval scale $[1, 5]$. Then for each of the twenty scenarios and for each group (symbolic vs. textual) we perform all pairwise comparisons between participant responses r_i and r_j, $i, j = 1 \ldots N, i \neq j$ to calculate the normalized distance $|r_i - r_j|/4$; the average of all these $N(N-1)/2$ distances is considered, N being the number of participants for each group. The resulting set consists of 2 (groups) × 20 (exercises) = 40 data points each expressing the level of total distance between the ratings of every pair of participants.

Figure 2 shows how these distances differ for various factors. In the upper graph, it can be observed that there is a difference between positive ("+","++") and negative ("−","−−") contributions, the latter yielding more disagreement. Probably more pronounced is the role of the satisfaction value of the origin goal (lower graph): satisfaction labels (**PS, FS**) lead to more agreement than denial labels (**PD, FD**) for both contribution signs.

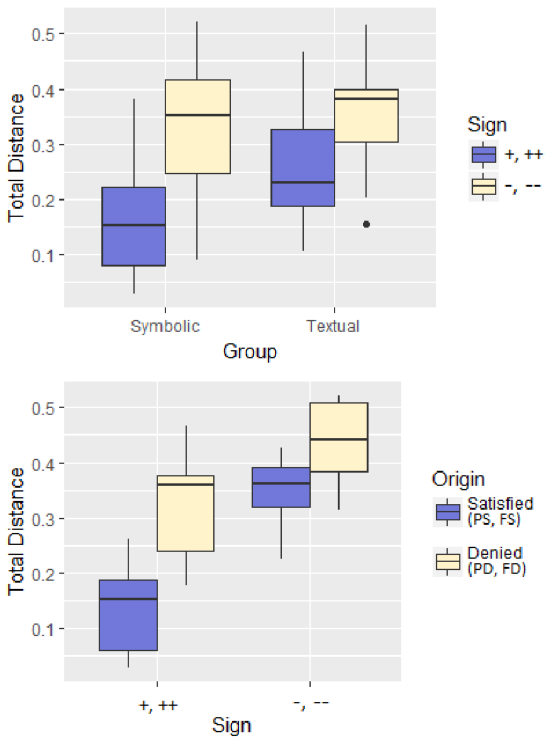

Figure 2. Disagreement wrt. Contribution Sign and Origin Satisfaction

3.2.4. Agreement with Normative Value. To measure the distance between the participant responses and the normative values according to the formal semantics (henceforth accuracy), we again coded both observations and normatives to the interval scale $[1, 5]$. For each single response obs_{ij} of participant i to exercise j with normative solution $norm_j$ the *relative* distance is $d_{ij} = obs_{ij} - norm_j$ and the *absolute* distance $|d_{ij}|$. Positive (rep. negative) relative distance implies that participants overestimate (resp. underestimate) satisfaction with respect to the normative. The relative distance, this time organized by contribution label, can be seen in Figure 3. In the figure, pp, p, n, nn stand for “++”,“+”,“−”,“−−” or *make*, *help*, *hurt*, *break* depending on the group considered. In the figure it can be observed that positive contribution labels lead to overestimation and negative ones to underestimation of the satisfaction of the destination. Less obvious in the figure is that, considering either absolute or relative distance, accuracy is higher in cases with positive contributions than in those with negative ones $F(1, 33) = 7.225, p < 0.05$ (absolute distance).

We further study the role of satisfaction status of the origin goal in accurately guessing the corresponding value for the destination goal. Ignoring cases where satisfaction of the origin goal is labeled as **N** (“No information”), our data show that satisfaction values lead to more accuracy than denial ones $F(1, 32) = 6.67, p < 0.05$, measured in absolute distance. Table 2 sheds some more light on this effect. The cells represent the average relative distance from normative. Focusing on cells where this is large (e.g. > 0.7) we can observe that (a) with *hurt* and *break* labels with denied origin goals the destination tends to be underestimated, (b) with *break* labels with satisfied origin goals satisfaction is overestimated and (c) with *make* links with denied origin goals, satisfaction of destination is overestimated. Given these, one can suspect that many participants do not seem to perceive the satisfaction inversion of negative labels that normative semantics suggest (observations (a) and (b) above). In addition, they do not seem to perceive that even a strong *makes* relationship can result in a fully denied destination goal, which is perfectly possible according to the normative semantics (observation (c) above). Further evidence to this is an interaction found between origin satisfaction and contribution sign $F(1, 32) = 8.3811, p < 0.01$ – considering relative distance. On one hand, for negative contributions, denial of the origin leads to underestimation of the destination satisfaction, i.e., some participants do not perceive inversion of denial to satisfaction. Positive contributions, on the other hand, appear to be perceived by some participants as difficult to cause full denial of the destination goal (they are positive contributions, after all), given the overestimation observed in such configurations.

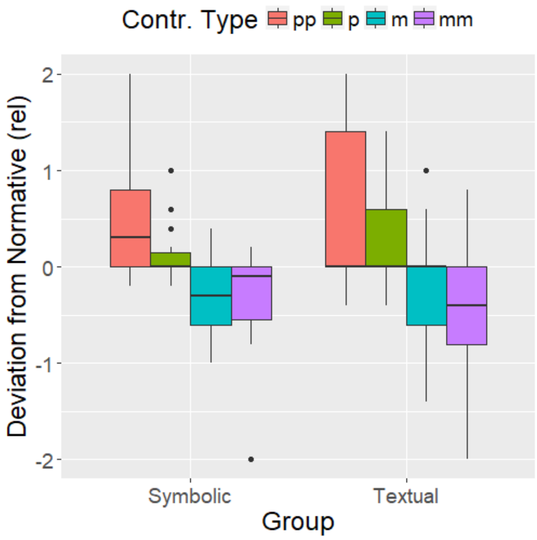

Figure 3. Relative Accuracy Level wrt. Contribution Label

TABLE 2. AVERAGE RELATIVE DISTANCE FROM NORMATIVE

Group	Origin	Make (++)	Help (+)	Hurt (-)	Break (–)
Symbolic	Sat	0.17	0.39	0.36	**0.72**
	Den	**0.83**	-0.14	**-1.03**	**-1.44**
Textual	Sat	-0.09	0.21	0.62	**0.82**
	Den	**1.41**	0.47	**-1.00**	**-1.88**

TABLE 3. AVERAGE RELATIVE DISTANCE FROM NORMATIVE FOR NO ORIGIN SATISFACTION

	Make (++)	Help (+)	Hurt (-)	Break (–)
Symbolic	0.44	0.22	-0.22	-0.44
Textual	0.41	0.12	-0.24	-0.35

Table 3 describes the average relative distance for only the cases in which there is "No information" **N** with regards to the satisfaction of the origin goal. According to the normative semantics all cells should be zero. However, participants assign satisfaction and denial to the destination goal of positive and negative contributions, respectively. It seems, then, that users perceive contributions as generators of satisfaction rather than just propagators thereof.

As a last note, neither visually nor statistically can we find any noteworthy effect in terms of response confidence. We omit graphs and analysis for the interest of space.

3.3. Discussion

3.3.1. Consequences. The results are of potential use for those wishing to apply goal modeling in practice, using current visualizations. There is firstly a distance between the normative semantics and what the participants understood that the construct means. Negative labels appear to associate with decreased accuracy. Negative labels also lead to more dispersion of the results, which may imply that the concept of negative contribution either is more exposed to alternative interpretations or is simply confusing. Further, participants appear to have problem comprehending denial values and how they interact with various labels. At the same time, participants seem to assign to labels the power to bring about satisfaction and denial irrespective of the satisfaction status of the origin goal. Finally, replacing labels with words seems to have no notable effect.

If these findings are confirmed by follow-up studies, they could interpret into useful advice for practicing goal modelers. For example, avoidance of negative contribution links may be preferred when possible, e.g., when modeling for decision exploration where relative rather than absolute contribution is important. In addition, in cases where conflict analysis is not necessary, it may also be preferable to avoid utilization of a denial variable.

3.3.2. Validity Threats. We now discuss external, internal, construct and statistical conclusion validity.

In terms of *external validity*, to appreciate the choice of our participant sample one must consider who is supposed to use goal models in practice. Although the common assumption is that languages like this are to be used by requirements analysts, we believe that such languages will reach their full potential when they are proved to also be usable by common stakeholders in requirements analysis contexts, i.e., the sources of requirements and ultimate decision makers. We assume that these are people drawn from a population that is a bit older and with more years of post-secondary education than our sample. Requirements analysts may also have great diversity of backgrounds. Regardless, we could not find empirical/survey data on the background of either analysts or stakeholders typically involved in requirements analysis problems. Nevertheless, the tasks performed by our participants – essentially interpreting symbols and words – do not presume any specific experience or training, mathematical or otherwise. As such we do not expect that a sample with more years of education or with longer field experience to yield different results, although this needs to be shown through follow-up studies.

At the same time, the models provided to the participants are isolated contribution links between two goals. In practice, inferences of satisfaction propagation within goal models take place in models with many goals and such contributions. The goals also have specific titles related to a domain, which was not the case in ours. Studying comprehension of links in the context of real goal diagrams and realistic inference and decision making problems would certainly offer more generalizability confidence, albeit introducing a host of additional variables to be controlled for. Such variables include perceptions of contribution link aggregation, model sizes and structures or modeled domains.

In terms of *internal validity*, the within-factors effects we identify can be claimed to be a result of biased video training: our training material for, e.g., negative contributions might not be as well done as that for the positive ones, hence the effect. While we are exploring ways to conclusively address such suspicion (e.g. use of independent judges), we believe it can be clarified through replications using alternative training instruments developed potentially by other researchers.

The topic of *construct validity* also requires some discussion. Recall that our main measures for "intuitiveness" are agreement (a) within participant responses and (b) between participant responses and language designers' expectations. These two measures of agreement strongly depend on the context in which they are measured, and in our case, as we saw, this is relatively narrow. Future studies may also consider these measures under different amounts of training, containing different degrees of explication of intended semantics or through more complex tasks such as assessing the outcome of more realistic and complex inferences that participants likely perform in reality.

On a final note, *statistical conclusion validity* may be a subject for further investigation given sensed departures from normality combined with effective sample sizes that are probably marginal in their ability to support a large sample size robustness argument [13]. Larger sample sizes may be needed as inferences to an entire population become more relevant.

4. Related Work

Many approaches for representing and automatically reasoning about propagation of goal satisfaction have been proposed in the goal oriented requirements engineering literature (e.g., [3], [15], [16], [8]). The criteria researchers use to evaluate their proposals are chiefly analytical, e.g., language expressiveness, amenability to useful and tractable automated reasoning or presence of a systematic elicitation approach.

At the same time, there have been several efforts to empirically investigate various measures of effectiveness of conceptual modeling languages. Common diagrammatic notations such as UML state diagrams and ER diagrams have, for example, been the subject of empirical investigation [17], [18]. In goal models, Horkoff et al. propose and evaluate an interactive evaluation technique for goal models [19]. Elsewhere, Hadar et al. [20] describe several studies in which goal diagrams and use case diagrams are compared on a variety of user tasks, such as reading and modification. Moody et al. investigate the appropriateness of visual choices of goal modeling languages such as *i** with respect to their comprehensibility, using established visualization rules [21]. This analytical work was followed-up by empirical work by Caire et al. [22] where visualization choices of the language are actually elicited from a crowd of experimental participants.

These efforts are indications that the research community is interested in the aspect of conceptual modeling that pertains to how visualization choices relate to how users perceive and use the models. Caire et al. for example [22] believe that users must play a primary role in defining how primitive constructs (e.g. goals and tasks) should be visualized. Our work demonstrates that we can go beyond just finding the right visualization for fixed semantic categories invented by language designers and approach the problem as one of agreement between what the constructs mean and how they are visualized, the difference being that both types of choices are open to influence from empirical evidence.

5. Summary and Future Work

We empirically tested the user perception of the meaning of the contribution link construct of a popular goal modeling language. We presented various such links to experimental participants and asked them to perform satisfaction propagation inference based on how the links looked like and some high-level training about what the links mean. We measured the agreement within their responses – by calculating average pair-wise distance – and the deviation of the responses from a normative value. We found that negative contributions and denial values lead to more response dispersion and deviation from normative semantics, while there was some difficulty to perceive the neutrality of propagation labels.

Apart from follow-up studies implied earlier (e.g., different samples of models and participants, different training material) and studies of specific interest to the goal modeling community, our main ambition for the future is to contribute to the body of experience in empirically assessing the effectiveness of conceptual modeling languages. Topics of interest include the consolidation of comprehensibility constructs and measures, the development of standardized assessment techniques (including self-reporting techniques) as well as investigation of the effects of individual differences in using conceptual models.

References

[1] E. S. K. Yu, "Towards modelling and reasoning support for early-phase requirements engineering," in *Proc. of the 3rd IEEE International Symposium on Requirements Engineering (RE'97)*, Annapolis, MD, 1997, pp. 226–235.

[2] A. Dardenne, A. van Lamsweerde, and S. Fickas, "Goal-directed requirements acquisition," *Science of Computer Programming*, vol. 20, no. 1-2, pp. 3–50, 1993.

[3] E. Letier and A. van Lamsweerde, "Reasoning about partial goal satisfaction for requirements and design engineering," in *Proc. of the 12th International Symposium on the Foundation of Software Engineering FSE-04*, Newport Beach, CA, 2004, pp. 53–62.

[4] J. Castro, M. Kolp, and J. Mylopoulos, "Towards requirements-driven information systems engineering: the Tropos project," *Information Systems*, vol. 27, no. 6, pp. 365–389, 2002.

[5] D. Amyot and G. Mussbacher, "User requirements notation: The first ten years, the next ten years," *Journal of Software (JSW)*, vol. 6, no. 5, pp. 747–768, 2011.

[6] D. Amyot, S. Ghanavati, J. Horkoff, G. Mussbacher, L. Peyton, and E. S. K. Yu, "Evaluating goal models within the goal-oriented requirement language," *International Journal of Intelligent Systems*, vol. 25, no. 8, pp. 841–877, 2010.

[7] J. Horkoff and E. Yu, "Finding solutions in goal models: An interactive backward reasoning approach," in *Proc. of the 29th International Conference on Conceptual Modeling (ER'10)*, Vancouver, BC, Canada, 2010, pp. 59–75.

[8] S. Liaskos, S. M. Khan, M. Soutchanski, and J. Mylopoulos, "Modeling and reasoning with decision-theoretic goals," in *Proc. of the 32th International Conference on Conceptual Modeling, (ER'13)*, Hong-Kong, China, 2013, pp. 19–32.

[9] P. Giorgini, J. Mylopoulos, E. Nicchiarelli, and R. Sebastiani, "Reasoning with goal models," in *Proc. of the 21st International Conf. on Conceptual Modeling (ER'02)*, London, UK, 2002, pp. 167–181.

[10] S. Liaskos, R. Jalman, and J. Aranda, "On eliciting preference and contribution measures in goal models," in *Proc. of the 20th International Requirements Engineering Conference (RE'12)*, Chicago, IL, 2012, pp. 221–230.

[11] P. Giorgini, J. Mylopoulos, E. Nicchiarelli, and R. Sebastiani, "Formal Reasoning Techniques for Goal Models". *Journal on Data Semantics I*, ser. LNCS, vol. 2800, pp. 1–20, 2010.

[12] D. Norman, *The Design of Everyday Things*. Basic Books, 2013.

[13] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 6th ed. Pearson, 2012.

[14] J. Fox, M. Friendly, and S. Weisberg, "Hypothesis tests for multivariate linear models using the car package," *R Journal*, vol. 5, no. 1, pp. 39–52, 2013.

[15] P. Giorgini, J. Mylopoulos, and R. Sebastiani, "Goal-oriented requirements analysis and reasoning in the Tropos methodology," *Engineering Applications of AI*, vol. 18, no. 2, pp. 159–171, 2005.

[16] A. van Lamsweerde, "Reasoning about alternative requirements options," in *Conceptual Modeling: Foundations and Applications*, ser. LNCS, vol. 5600, 2009, pp. 380–397.

[17] J. A. Cruz-Lemus, M. Genero, M. E. Manso, S. Morasca, and M. Piattini, "Assessing the understandability of UML statechart diagrams with composite states—a family of empirical studies," *Empirical Software Engineering*, vol. 14, no. 6, pp. 685–719, 2009.

[18] H. C. Purchase, R. Welland, M. McGill, and L. Colpoys, "Comprehension of diagram syntax: an empirical study of entity relationship notations," *International Journal of Human-Computer Studies*, vol. 61, no. 2, pp. 187 – 203, 2004.

[19] J. Horkoff and E. S. K. Yu, "Interactive goal model analysis for early requirements engineering," *Requirements Engineering*, vol. 21, no. 1, pp. 29–61, 2016.

[20] I. Hadar, I. Reinhartz-Berger, T. Kuflik, A. Perini, F. Ricca, and A. Susi, "Comparing the comprehensibility of requirements models expressed in use case and Tropos: Results from a family of experiments," *Information and Software Technology*, vol. 55, no. 10, pp. 1823 – 1843, 2013.

[21] D. Moody, "The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering," *IEEE Trans. on Software Engineering*, vol. 35, no. 6, pp. 756 – 779, 2009.

[22] P. Caire, N. Genon, P. Heymans, and D. L. Moody, "Visual notation design 2.0: Towards user comprehensible requirements engineering notations," in *Proc. of the 21st IEEE International Requirements Engineering Conference (RE'13)*, Rio de Janeiro, Brazil, 2013, pp. 115–124.