Preliminary Results on the Understandability of i^* Notations

Ralf Laue[1], Frank Hogrebe[2], Boris Böttcher[3], and Markus Nüttgens[3]

[1] University of Applied Sciences of Zwickau, Germany
Department of Information Science
ralf.laue@fh-zwickau.de
[2] University of Applied Sciences
Hessische Hochschule für Polizei und Verwaltung Wiesbaden, Germany
Department of Public Administration
frank.hogrebe@hfpv-hessen.de
[3] University of Hamburg, Germany
boris.boettcher@uni-hamburg.de,markus.nuettgens@wiso.uni-hamburg.de

Abstract. In this article, we describe the first results of two series of experiments. It is the aim of our experiments to compare the performance of novice users working with i^* models. We compared the results achieved with i^* diagrams in the traditional notation with the results using the alternative notation which has been suggested by using Moody's "Physics of Notations" framework. We created two models in both variants of the notation. In the first experiment, we asked comprehension questions about models printed on paper. In a second experiment, the participants had to answer questions using models on a computer screen. The results give first support to the hypothesis that the improvements suggested by Moody could be helpful for understanding i^* models. We also identified a detail for which the newly suggested notation variant can lead to understandability problems and should be improved.

1 Introduction

The language i^* [1] is one of the most prominent visual languages in the field of requirements engineering. The graphical elements used in this notation have been introduced in Yu's seminal work without discussing the question, why a certain notation element (shape, line, arrow etc.) should be used for expressing a certain concept. Moody [2] stated that this lack of discussion about the visual syntax is typical for visual languages used in computer science. To overcome this situation, he presented the "Physics of Notations" (PoN), a theoretically well-founded framework for evaluating the quality of visual notations. The framework can also be used for designing cognitively effective visual notations. In [3, 4], the i^* modelling language has been discussed according to the PoN framework. Shortcomings of the currently used notation have been found, and an alternative visual syntax for i^* that follows the PoN principles has been suggested.

In particular, Moody and his co-authors suggest to replace the shapes traditionally used in i^* (see Fig. 1) by another set of graphical symbols (see Fig. 2).

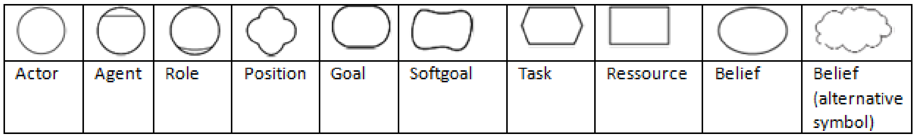


Fig. 1. traditional i^* symbols

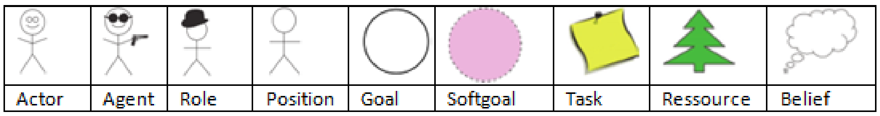


Fig. 2. i^* symbols as suggested in [3, 4]

We will refer to this notation as PoN notation. When comparing Fig. 1 and Fig. 2, we see that it is easier to differentiate between the symbols in the PoN notation. Furthermore, it should be easier to understand the meaning of these symbols by intuition. For example the concept of a "Task" is symbolized by a stick-notice which is more intuitive than the traditional hexagon symbol.

Several arguments of the PoN framework are backed by empirical validation on understandability of symbols used in graphical user interfaces: Byrne [5] showed that simply designed icons can be recognized easier. Goonetilleke et al. [6] found that icons can be used better if they clearly differ from each other.

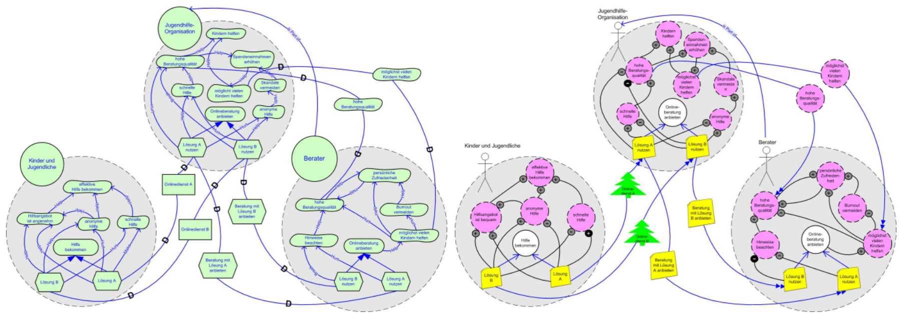
However, there are no empirical results for evaluating the improvements suggested in [3, 4] as a whole. The authors of these articles explicitly mention that their recommendations for improving the notation have not been empirically tested and call for this kind of empirical work. In this paper, we describe some first experiments to find out whether users of i^* models perform better when using the improvements of the visual syntax that have been suggested by Moody et al.

2 Experiment 1: Working with a Model on Paper

2.1 Experimental Setup

For two scenarios, we constructed two i^* models: One using the traditional i^* notation and another one using the notation suggested by Moody et al. The first scenario was the classic meeting scheduler problem [1], the other one was the youth counseling example from [7].

Students of public management and students of information science were asked to answer comprehension questions about those models. For this purpose, all students got an introduction lecture on the concepts of i^*, but without referring to a specific visual notation. Before the comprehension experiment, they have been assigned to two homogeneous groups - one for each notation variant. An introduction of the i^* symbols (the traditional ones in one group, the PoN



(a) traditional notation (b) PoN notation

Fig. 3. Two variants of a model

symbols in the other one) was given to the students before they had to work with the models (printed in colour on paper) in order to answer 6 paper-based comprehension questions. While working with the models, each student had a legend explaining all the symbols in form of a one-page handout. No time limit was given for answering the questions.

The purpose of three questions was to reason about the contribution that model elements have to softgoals. An example question would be “Which softgoal of the youth counseling organization will be affected negatively when solution A is used?” One question was aimed to measure understanding of task decomposition. Two questions were “counting questions” (such as: How many actors depend on the appointment participant?)

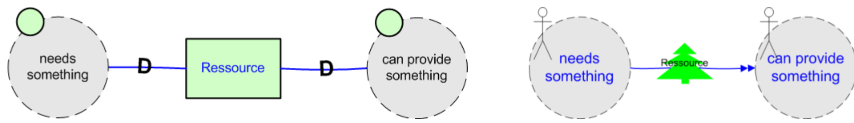
Altogether, the questions have been answered by 121 persons so far (32 bachelor students of public management, 53 master students of public management, 36 bachelor students of information science).

2.2 Results

Tab. 1 shows how well the participants answered the questions. Our assumption that the users who worked with the PoN models would perform better could not be confirmed. In one group (bachelor students of public management) the participants using the PoN notation answered more questions wrongly than the participants working with the traditional notation. After this unexpected result, we had a closer look at the results of each question for this group of participants. We identified one “counting” question for which in this group of students,

	traditional notation	PoN notation
Public Management, BA	56.8	53.3
Public Management, MA	47.5	55.7
Information Science, BA	54.2	60.4

Table 1. Percentage of correctly answered questions (all questions)



(a) traditional notation (b) PoN notation

Fig. 4. Notations for expressing a dependency

the participants using the PoN notation had less than half as much correct answers as the participants using the traditional notation. This question required to answer how many actors depend on a given actor. For solving it correctly, it was important to understand the *direction* of the dependency (Who provides something and who needs something?) As shown in Fig. 4, it is easy to misunderstand the PoN notation to depict these facts: An intuitive interpretation of the arc symbol could be that the resource is transported from the origin of the arc in the direction of the arrowhead (comparable to the intuitive interpretation of causal relationships reported in [8]). However, the concept of "dependency" is defined in the opposite direction. Using the traditional "D" symbol to depict a dependency avoids this kind of misunderstandings. Therefore we think that the results of our experiments are an indicator that the notation for dependencies in the PoN notation should be improved (see Fig. 4).

If we exclude the question that requires to count dependencies, the results are as shown in Tab. 2. While the results shown in this table show a trend in the expected direction, they are not statistically significant, and experiment 1 does not confirm the assumption that working with models in the PoN notation leads to less errors. In the future, we plan additional experiments in order to get results which are founded on the performance of a larger number of participants.

	traditional notation	PoN notation
Public Management, BA	49.0	50.0
Public Management, MA	43.1	48.0
Information Science, BA	48.5	55.5

Table 2. Percentage of correctly answered questions (dependency counting question removed)

3 Experiment 2: Working with a Model on a Computer Screen

3.1 Experimental Setup

In a second experiment, we used the eye tracking device Tobii T60 XL Eye Tracker with a 24 in. screen and the software Tobii Studio 3.2.1. Sixteen participants (eight for each notation variant) had to solve a set of six tasks while their eye movement was recorded. There were 3 students of business information science in the group using the traditional notation and 4 students of business information science in the group using the PoN notation. All other participants were recruited from the academic staff at the department of business information science at the University of Hamburg.

Before the experiment, each participant got an introduction to i^* by means of a self-learning course (they were not familiar with i^* before). Two different variants of the course were developed. One used the traditional i^* notation, the second one the PoN notation. In all other aspects, the self-learning courses were identical. The participants had as much time as they wished to get familiar with the i^* concepts and the notation, and they had the possibility to ask additional questions if something remained unclear.

Then the participants had to solve six questions on the same i^* diagrams that have been used in experiment 1. We could not always use the same questions as in experiment 1, because in experiment 2 the participants had to remember the questions. Therefore, some questions had to be simplified or changed. The participants were instructed to press a key as quickly as possible if they have found the answer. Afterwards, they showed their answer on the screen. This way, we recorded both the time needed to find the answer as the eye movement of the participants while looking at the diagrams in order to answer a question.

3.2 Results

As already observed in experiment 1, the correctness of the answers to a question where the number of dependencies had to be given was better in the group using the traditional notation (4 errors, compared to 5 errors in the PoN group). Tab. 3 shows the number of wrongly answered questions for all participants.

If we include all six questions, the mean number of errors in the group using the traditional notation is 2.625. For the PoN group, it is 1.125. We performed a one-tailed Mann-Whitney U-test on the data groups. Both sample sizes were 8, the U value for the one-tailed Mann-Whitney test is 11 and the P-value is 0.014, i.e. the result is significant at the standard level of 95%. If we exclude the "dependency counting" question, we even get a highly significant result (P-value for the one-tailed test 0.00148).

A large amount of data has been collected by the eye-tracking system. We expect additional insights on how people work with i^* models by analyzing this data, but have not yet started the analysis.

	trad. notation								PoN notation							
Errors (all six questions)	3	2	2	3	2	2	5	2	0	1	1	0	2	1	2	3
Errors (without the "dependency counting" question)	2	1	2	2	2	2	4	2	0	1	0	0	1	0	1	2

Table 3. Wrongly answered questions by the participants of experiment 2

4 Discussion and Conclusions

This paper describes the first preliminary results of two experiments. Experiment 1 could not confirm in a statistically significant way that working with the PoN notation has a positive effect on understanding the diagrams. We plan to repeat the experiments with more participants in order to get more reliable results. However, one side-result of the experiment (which has been backed by experiment 2 as well) was the suggestion to replace the depiction of the dependency link in the PoN notation. Experiment 2 showed that participants using the PoN notation performed significantly better (working on a computer screen). The statistical significance of this result is very high despite the fact that the number of participants was rather low (which is not uncommon for eye-tracking experiments).

A possible reason for the high significance achieved in experiment 2 is that in the eye-tracking environment, the participants had to rely on their memory for deciding about the meaning of a graphical symbol - possibly a situation where easy-to-remember symbols are particularly helpful.

For the future, we plan to repeat our experiment with more participants. Furthermore, we expect additional insights by analysing the logs of the eye-tracking sessions.

A limitation of our research is that all experiments have been conducted with novice users, therefore the results cannot be transferred to experienced i^* users.

References

1. Yu, E.S.K.: Towards modeling and reasoning support for early-phase requirements engineering. In: 3rd IEEE International Symposium on Requirements Engineering. (1997) 226–235
2. Moody, D.L.: The "physics" of notations: a scientific approach to designing visual notations in software engineering. In: 32nd ACM/IEEE International Conference on Software Engineering. (2010) 485–486
3. Moody, D.L., Heymans, P., Matulevicius, R.: Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation. Requir. Eng. **15** (2010) 141–175
4. Moody, D.L., Heymans, P., Matulevicius, R.: Improving the effectiveness of visual representations in requirements engineering: An evaluation of i* visual syntax. In: 17th IEEE International Requirements Engineering Conference. (2009) 171–180
5. Byrne, M.D.: Using icons to find documents: simplicity is critical. In: Human-Computer Interaction, INTERACT '93. (1993) 446–453
6. Goonetilleke, R.S., Shih, H.M., On, H.K., Fritsch, J.: Effects of training and representational characteristics in icon design. Int. J. Hum.-Comput. Stud. **55** (2001) 741–760
7. Horkoff, J., Yu, E.S.K.: Evaluating goal achievement in enterprise modeling - an interactive procedure and experiences. In: The Practice of Enterprise Modeling, Working Conference. Volume 39 of LNBIP., Springer (2009) 145–160
8. Winn, W.: Encoding and retrieval of information in maps and diagrams. IEEE Transactions on Professional Communication **33** (1990) 103–107