Increasing the Semantic Transparency of the KAOS Goal Model Concrete Syntax

Mafalda Santos, Catarina Gralha, Miguel Goulão, and João Araújo

NOVA LINCS, DI, FCT, Universidade NOVA de Lisboa
{mcd.santos, acg.almeida}@campus.fct.unl.pt,
{mgoul, joao.araujo}@fct.unl.pt

Abstract. Stakeholders without formal training in requirements modelling languages, such as KAOS, struggle to understand requirements specifications. The lack of semantic transparency of the KAOS goal model concrete syntax is perceived as a communication barrier between stakeholders and requirements engineers. We report on a series of related empirical experiments that include the proposal of alternative concrete syntaxes for KAOS by leveraging design contributions from novices and their evaluation with respect to semantic transparency, in contrast with the standard KAOS goal model concrete syntax. We propose an alternative concrete syntax for KAOS that increases its semantic transparency (mean difference of .23, in [-1.00..1.00]) leading to a significantly higher correct symbol identification (mean difference of 19%) by novices. These results may be a stepping stone for reducing the communication gap between stakeholders and requirements engineers.

Keywords: goal model concrete syntax, empirical evaluation, KAOS

1 Introduction

Goal-oriented requirements models [1] are important for requirements elicitation and analysis, where communication with stakeholders plays a major role. For this to be effective, both requirements engineers and other stakeholders must have a common understanding of these models to participate in their development process. Yet, most of the stakeholders have no previous knowledge about requirements modelling languages, requiring formal training on such models. Requirements modelling languages themselves are under scrutiny, with proposals for their evolution, which include syntactic and semantic aspects. However, most of the research on these languages focuses on semantic aspects, rather than syntactic ones. Neglecting semantic transparency raises a communication barrier between requirements engineers and stakeholders. Semantic transparency has a positive impact on cognitive effectiveness, which is based on the ease, speed, and accuracy to process the information illustrated on a requirements model.

We are interested in evaluating the semantic transparency of the concrete syntaxes of goal-oriented requirements modelling languages, in particular, the KAOS goal model. We conducted a series of quasi-experiments that evaluate the cognitive effectiveness of KAOS concrete syntax. The approach followed is based on the "Physics" of Notations (PoN) [2]. We present an alternative concrete syntax for KAOS goal models proposed by novices. We evaluate it with respect to semantic transparency, in contrast with the standard KAOS goal model concrete syntax. The results suggest that the alternative concrete syntax improves cognitive effectiveness as it significantly increases semantic transparency, leading to a higher correct symbol identification by novices.

The paper is structured as follows. Section 2 discusses KAOS, PoN, and related work. Sections 4 to 7 describe the 4 studies that were undertaken to evaluate the concrete syntax of the KAOS goal model. Section 8 discusses the results and Section 9 draws some conclusions and discusses future work.

2 Background

The KAOS methodology. Among the main Goal-Oriented Requirements Engineering (GORE) approaches [1, 4–9], KAOS has been one of the most relevant. Its emphasis is on semi-formal and formal reasoning about behavioural goals to derive goal refinements, operationalisations, conflict management and risk analysis [10]. Goals are a prescriptive intention statement about a system whose satisfaction, in general, needs cooperation of agents that configure the system. Through *and/or* decompositions, goals can be refined into subgoals, requirements or expectations. Objects can be specified to describe the project structural model. Obstacles and goals relations can be used to identify system vulnerabilities [11]. In Fig. 1 we show 18 of the KAOS concepts that we use here.

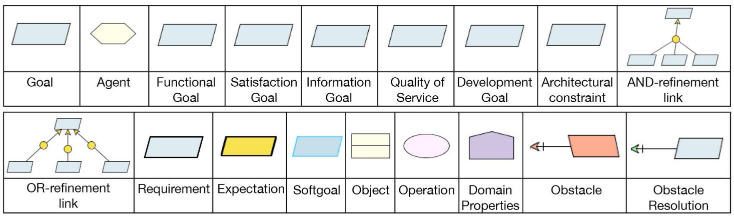


Fig. 1: Standard KAOS symbol set [1, 4]

The "Physics" of Notations. Visual notations are used for communication among different kinds of stakeholders (e.g., developers, end-users, clients), where cognitive effectiveness is key to support design and problem solving. Improving the cognitive effectiveness of visual notations will promote their use and allow different stakeholders to share the same understanding of the software. The PoN theory [2] consists of a set of principles for designing cognitively effective visual notations, *optimised for processing by the human mind* [2]. These principles are related to several aspects of a modelling language such as semantic transparency, complexity management, cognitive integration, cognitive adjustment, graphic economy, semiotic clarity, visual expressiveness, double codification and perceptual discriminability. Our work focuses on the principle of *semantic transparency*, which argues that the symbols of a language must suggest their meaning: there must be a correspondence between the visual properties and the semantic properties of the represented objects [12]. Based on the notion of natural mapping, [13], a graphic language should take advantage of physical analogies, visual metaphors, common logical properties and cultural associations. Semantically transparent symbols allow to reduce the cognitive load in their recognition process as they use visual mnemonics, so their meaning can be easily deduced.

Related work. Caire et al. [3] applied the PoN theory to *i**, and analysed the principle of semantic transparency by involving novices in the design of a new *i** concrete syntax. The results reveal that novices have a better performance than experts in creating symbols for *i**. The new symbols improved the semantic transparency compared to the standard *i** concrete syntax, suggesting that visual notations should be designed by novices, not experts. Here, we replicated this approach to KAOS, finding similar results. Our work differs from Caire et al.'s as we categorise participants (with or without knowledge on modelling languages), and include the use of the created symbols to model a small problem.

Other studies applied the PoN theory. In their seminal work, Moody et al. [25] applied the PoN principles to evaluate *i**. The authors identified flaws in the *i** concrete syntax and proposed improvements. Genon et al. [14] applied PoN to Use Case Maps. The authors analysed all the PoN principles themselves, neither proposing an alternative concrete sytax nor involving novices in the process, as we do in this paper. For each principle, they identified weaknesses and suggested improvements. The same authors identically applied the PoN to BPMN [15]. Granada et al. [16] applied the PoN to WebML. They analysed each PoN principle themselves, identifying weaknesses and proposing improvements. For the semantic transparency principle, they performed an empirical study with two groups: postgraduate students and experts in software engineering. The results showed that some WebML symbols have semantic transparency problems. Also, their results suggest that experts have a greater ability to infer the meaning from the appearance of the symbol. Saleh et al. [28] applied the PoN to misuse cases. The authors analysed all the PoN principles themselves, and proposed a new concrete syntax for misuse cases. They then compared the cognitive effectiveness of the original and the new concrete syntaxes. The results indicated that the new concrete syntax is more semantically transparent than the original one.

Matulevičius et al. [26] evaluated how KAOS and its supporting tool (Objectiver), help modellers to adhere to 9 visual modelling principles, during the modelling activity. The authors offered recommendations for modellers, language engineers and tool developers. The same authors evaluated the quality of *i** and KAOS [27]. The evaluation consisted of interviews, goal models creation, evaluation of the models and the modelling languages. The results revealed that the semantics of these languages is not defined clearly enough. Finally, Boone et al. [29] applied the PoN to CHOOSE. The authors evaluated 5 PoN principles, and created 3 alternative concrete syntax based on the results of their initial evaluation. All the concrete syntaxes were evaluated throughout an empirical study with business engineering students. One the new concrete syntaxes outperformed the others in terms of cognitive effectiveness.

3 Research planning

Research questions. Three research questions guided our quasi-experiments on the semantic transparency of KAOS goal models:

RQ1. Is the KAOS visual notation semantically opaque?

RQ2. Can participants with no knowledge in modelling languages design more semantically transparent symbols than participants with knowledge in modelling languages?

RQ3. Which visual notation *(standard, stereotype, or prototype)* is more semantically transparent?

Research design. The research design consists of 4 related empirical studies, where the results of the earlier studies provide inputs to the later studies.

1. **Symbolisation experiment**: a group of 99 novice participants designed symbols for KAOS concepts, a task normally reserved for experts;
2. **Stereotyping analysis**: we identified and organised categories with the most common symbols produced for each KAOS concept. This defined the *stereotype symbol set*.
3. **Prototyping experiment**: a group of 88 novice-participants chose the symbols they consider to better represent each KAOS concept. The most voted symbols for each KAOS concept defined the *prototype symbol set*.
4. **Semantic transparency experiment**: we evaluated the ability of 52 participants to infer the meanings of novice-designed symbols (*stereotype* and *prototype symbol set*) compared to expert-designed symbols (*standard KAOS*).

For studies 1, 3 and 4, we used questionnaires, which are explained in the next sections. These questionnaires can be found at `https://goo.gl/G1aDkg`.

4 Study 1 – Symbolisation experiment

Goals. The goal of this study was to **obtain candidate symbols drawn by novices to illustrate 18 KAOS goal models concepts**. We used the sign production technique [17]. This involves asking members of the target audience (i.e., those who will be interpreting the models) to generate symbols to represent a set of given concepts. The rationale is that symbols produced by members of the target audience are more likely to be understood and recognised by other members of the target audience, due to their common cognitive profile [17, 18]. This approach has been used to design public information symbols [19], office equipment symbols [20], workflow modelling [21], icons for graphical user interfaces [18], and to design RE visual notations, such as *i** [3]. This type of studies has consistently shown that symbols produced in this way are more accurately interpreted by their target audience than those produced by experts.

Participants. There were 99 participants (73 males, 26 females), all students from Universidade Nova de Lisboa (UNL), from different courses (Mechanical Engineering, Industrial Engineering and Management, Environmental Engineering, Civil Engineering, and Computer Science). This diversity was deliberate, as we wanted our participants to be surrogates of stakeholders from different backgrounds who will interact with requirements engineers. We categorised the participants in: **W**ith **N**o **K**nowledge in **M**odelling **L**anguages (**WNKML**) and **W**ith **K**nowledge in **M**odelling **L**anguages (**WKML**). Participants were recruited through convenience sampling and participated voluntarily. The **WNKML** group had 53 participants (32 males, 21 females; 53 undergraduates), from courses other than Computer Science. They had no previous knowledge of modelling languages in general, or KAOS in particular, and represent stakeholders from other domains in our study. The **WKML** group had 46 participants (41 males, 5 females; 40 undergraduates, 6 MSc students), all Computer Science students. All of them had previous knowledge of modelling languages, and 37 had a brief contact with KAOS, in the context of a Software Engineering course. They are representatives of stakeholders with some technical background, but no expertise in Requirements Engineering goal models.

Experimental material. Each participant was provided with a 6-pages questionnaire and a pen or pencil. The first page had the instructions for answering the questionnaire, which is divided into 3 parts: *Part I* provided the definition of 18 KAOS concepts. For each concept, participants were asked to create a visual representation inside a framed area, to constrain the size of the proposed symbols. *Part II* provided a requirements description. We asked participants to represent it using the visual symbols they proposed in *Part I*. *Part III* was used for collecting demographic data on participants.

Procedure. Participants were instructed to produce drawings expressing the meaning of each concept. None of the questions was mandatory. No time limit was set but, on average, participants took 45-60 minutes to complete the tasks.

Results. The participants produced a total of 1518 symbols, 723 of which were drawn by the WNKML and 795 by the WKML group. This corresponds to a response rate of 85.2%. There is no deviation from normality according to the Kolmogorov-Smirnov and the Shapiro-Wilk tests (p>.05), suggesting the response rate difference is normally distributed. There were no outliers. We conducted a paired-samples t-test to compare the response rates for each concept, from WKML and WNKML participants. This test was found to be statistically significant, $t(17)$=-8.135, p<.001; d=2.278. The effect size for this analysis (d=2.378) was found to exceed Cohen's convention for a large effect (d=.80). These results suggest that participants from the WKML group (M=.960, SD=.033) had a higher response rate than participants from the WNKML group (M=.758, SD=.125).

The response rate for constructing a requirements model, using a requirements description provided with the questionnaire, was considerably lower (68.6%) than the response rate for symbol proposals. The WKML group had a higher response rate (97.8%) than the WNKML group (43.4%). These results suggest that the WNKML group had more difficulty in building the KAOS model. The overall results suggest that both groups encountered more difficulties when creating the KAOS model than when proposing symbols for each concept. The WNKML group had more difficulty than the WKML, in both parts of the questionnaire.

5 Study 2 – Stereotyping analysis

Goals. The goal of this study was to identify **the most common symbols produced by the participants** – the *stereotype symbol set* – for each KAOS concept in Study 1 (Section 4). This approach is based on the assumption that the most frequently produced symbol for representing a concept is also the most frequently recognised by the members of the target audience [17, 18].

Procedure. The symbols produced in Study 1 were classified into symbol categories. Each symbol category represents all the symbols containing common visual features (e.g. the smiling face in Fig. 2 represents a category of similar drawings to convey the concept of *Agent*). Participants from the WKML group seem to have been influenced by the modelling languages they know. Their symbols, for each concept, were less varied, more conventional and more abstract. In contrast, the symbols created by the WNKML group were more detailed and creative, being rather varied among the participants. For both groups, we categorised the symbols based on their visual and conceptual similarity. We then combined the categories of symbols produced by both groups (naturally, some of them were the same) and counted the number of members in each category. We then selected the most representative category (i.e. the one with the highest number of members) for each concept, resulting in the *stereotype symbol set.*

Results. This study resulted in a symbol categories table, containing all the symbols produced by the participants WKML and WNKML. We extracted from it the *stereotype symbol set* (Fig. 2), with the most common symbols produced for each of the 18 KAOS concepts. The degree of stereotypy [17], or stereotype weight [18], measures the level of consensus about a concept visual representation. The average degree of stereotypy of the stereotype symbols was .212% (*SD*=.128). *Agent* is the only outlier (and the maximum value), with a stereotypy of .660. Indeed, *Agent* was the only concept with a majority, i.e., a degree of stereotypy above .5. These results confirm the inherent difficulty in representing such abstract concepts [18]. A paired samples t-test indicated that the degree of stereotypy is not significantly different for the WKML group (M=.224, SD=.148) and the WNKML group (M=.200, SD=.151), $t(17)$=.637, p=.533, suggesting both groups contributed similarly to the *stereotype symbol set.*

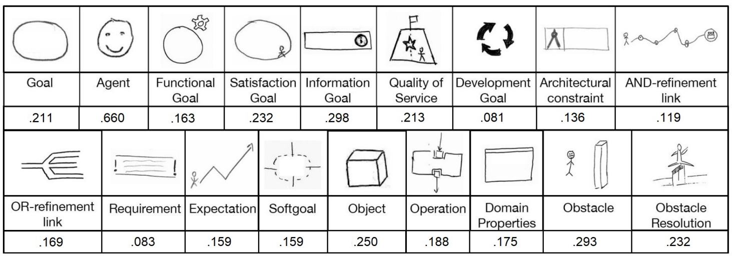


Fig. 2: KAOS stereotype symbol set with symbols degree of stereotypy

6 Study 3 – Prototyping experiment

Goals. Stereotyping can been challenged on the grounds that the most frequently produced drawings may not necessarily be the ones that convey concepts most effectively. In fact, visual metaphors often work well as a mnemonic for the concept name, while failing to represent the concept itself [3, 18]. In this study, we asked novice-participants to analyse symbols produced in Study 1 (Section 4) and **choose which best represents each KAOS concept**. The most frequently chosen symbol for each concept was then included in the *prototype symbol set*. This represents a consensual perception by members of the target audience about semantic transparency.

Participants. There were 80 participants (70 males and 10 females), all students from UNL. None of them had participated in Study 1, to prevent bias concerning their own proposals. Again, we categorised the participants into the WNKML and WKML groups. All students were recruited through convenience sampling and participated voluntarily. The WNKML group had 56 participants (48 males, 8 females; 56 undergraduates), all undergraduate students in Computer Science or engineering: 43 from Computer Science, 13 from other Engineering courses (Mechanical Engineering, Industrial Engineering and Management and Civil Engineering). The WKML group had 24 participants (22 males, 2 females; 8 undergraduates, 16 MSc students).

Experimental material. Each participant was provided with a 4-page questionnaire and a pen or pencil. The questionnaire was divided into 2 parts. *Part I* provided 18 KAOS goal model concept descriptions, each with a corresponding set containing from 3 to 7 candidate symbols. The candidate symbols represented the categories corresponding to that KAOS concept with the highest stereotypy. Participants were asked to choose the symbol that represents the best visual metaphor for each concept. In *Part II*, we collected demographic data.

Procedure. Participants were verbally instructed to answer the questionnaire, by choosing the symbols that, in their opinion, best expressed the meaning of the KAOS concepts. None of the questions was mandatory. No time limit was set but, on average, participants took 5-20 minutes to complete the questionnaire.

Results. The result of this study is the *prototype symbol set* (Fig. 3), composed by the most voted symbol to represent each one of the 18 KAOS concepts. The overall level of consensus among judgement was lower than .5 for most symbols. Only 5 were selected by more than half of the participants. The Koklmogorov-Smirnov and Shapiro-Wilk normality tests were used to assess the normality of the distribution of the number of elements selected by each participant that made it into the prototype, for WNKML and WKML groups. The distribution among members of the WKML group departed from normality (p=<.022 and p=.096, respectively). We conducted a Welch's t-test, $t(19.442)$=9.735, p=.006 to evaluate whether there were any significant differences between the number of elements selected by the WKML group (M=6.25, SD=3.357) and by the WNKML group (M=9.05, SD=2.522). On average, participants from the WKML group selected less voted elements than those from the WNKML.

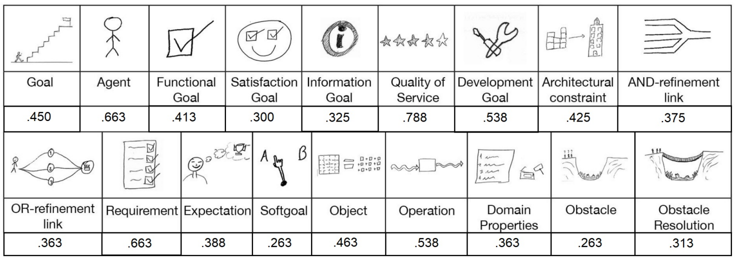


Fig. 3: KAOS prototype symbol set with level of consensus

7 Study 4 – Semantic transparency experiment

Goals. The goal of this study was to **evaluate the semantic transparency of 3 symbol sets for KAOS**: *standard* (Fig. 1), *stereotype* (Fig. 2) and *prototype* (Fig. 3). Semantic transparency defines the degree of association between a symbol's form and content [2]. We conducted a blind interpretation study where participants infered the concept (content) associated with each symbol (form). This method is recommended by the International Organization for Standardization (ISO) *for testing the comprehensibility of graphical symbols* [22]. ISO uses it when testing the comprehensibility of standard symbols prior to their release, which can be measured by the percentage of correct responses (i.e., hit rate).
Participants. There were 52 participants (44 males, 8 females), all students from UNL except 2 students from Universidade de Lisboa. We used different participants than in Study 1 and Study 3. We categorised the participants into the WNKML and WKML groups. All students were recruited through convenience sampling and participated voluntarily. The WNKML group had 17 participants (15 males, 2 females; 14 undergraduates, 3 MSc students) from Computer Science, Mechanical Engineering, Industrial Engineering and Management and Architecture. The WKML group had 35 participants (29 males, 6 females; 5 undergraduates, 30 MSc students), all Computer Science students.
Experimental material. Each participant was provided with a 5-page questionnaire and a pen or pencil. The questionnaire was divided into 4 parts. In *Part I*, we provide 18 KAOS concepts and descriptions. In *Part II*, participants are asked to fill a Matching Table, by matching the symbols from each of the 3 symbol sets with each of the 18 KAOS concepts. In *Part III*, we provide a table containing the 3 symbol sets. Finally, in *Part IV* we collected demographic data.
Procedure. Participants were instructed verbally to answer the questionnaire, by selecting a symbol from each set that better described each KAOS concept. None of the questions was mandatory. No time limit was set but, on average, participants took 20-50 minutes do complete the questionnaire.
Hypotheses, parameters and variables. The independent variable is the *symbol set* (i.e., *standard*, *stereotype* or *prototype*). The dependent variables are the *semantic transparency coefficient* [2, 3], the degree of proximity between a symbol and the semantic construct represented by it; and the *hit rate*, an

indicator for measuring symbols comprehension. For each one of the dependent variables, we have defined 3 hypothesis, which we present in Table 1.

Table 1: Hypotheses for Semantic Transparency and Hit Rate

Hypotheses	**Description**
H_{1ST}	Stereotype is more semantically transparent than standard
H_{2ST}	Prototype is more semantically transparent than standard
H_{3ST}	Prototype is more semantically transparent than stereotype
H_{4HR}	Stereotype has a higher hit rate than standard
H_{5HR}	Prototype has a higher hit rate than standard
H_{6HR}	Prototype has a higher hit rate than stereotype

We predict that the stereotype and prototype symbol sets would outperform the standard KAOS. We also predict that the prototype would outperform the stereotype symbol set, since we consider that the most chosen symbols are easier to interpret than the most common ones, resulting in the following ordering:

prototype symbol set > *stereotype symbol set* > *standard symbol set*

Results. Table 2 shows the semantic transparency coefficient and hit rate for the 3 symbol sets. A symbol's semantic transparency coefficient is given by [3]: $\frac{maximum\ frequency - expected\ frequency}{total\ responses - expected\ frequency}$. For the semantic transparency coefficient, each cell is coloured from red (semantically opaque symbol) to green (semantically transparent symbol). For the hit rate, the highlighted values correspond to the symbols that respect the ISO threshold for comprehensibility ($\geq 67\%$) [23].

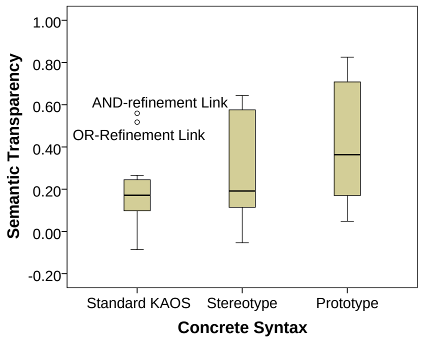
Table 2: Semantic Transparency coefficient and Hit Rate results

Symbol	**Semant. Transp. coefficient**			**Hit Rate**		
	standard	stereotype	prototype	standard	stereotype	prototype
Agent	0.20	0.58	**0.81**	26.9%	59.9%	**81.9%**
AND-refinement link	**0.56**	0.35	0.39	**59.6**%	38.5%	42.2%
Archit. constraint	-0.09	0.05	**0.17**	0.5%	10.1%	**21.6%**
Development goal	0.01	**0.19**	0.17	9.6%	**23.1** %	21.2%
Domain properties	0.16	-0.05	**0.23**	23.1%	0.5%	**26.9**%
Expectation	0.16	0.20	**0.44**	23.1%	24.2%	**47.3%**
Functional goal	0.16	**0.17**	0.14	**23.1%**	21.8%	18.3%
Goal	0.20	0.11	**0.24**	26.9%	16.3%	**28.1**%
Information goal	0.18	0.63	**0.75**	25.0%	65.1%	**76.0%**
Object	0.10	**0.60**	0.15	17.3%	**62.6%**	19.7%
Obstacle	0.27	0.64	**0.83**	32.7%	66.3%	**83.5%**
Obstacle resolution	0.24	0.60	**0.76**	30.8%	62.5%	**77.6%**
Operation	0.26	0.41	**0.44**	31.7%	43.9%	**46.8%**
OR-refinement link	**0.52**	0.42	0.34	**55.8%**	45.2%	37.6%
Quality of service	-0.03	0.11	**0.63**	5.8%	16.3%	**64.6%**
Requirement	**0.22**	0.04	0.18	**28.8%**	9.6%	22.2%
Satisfaction Goal	0.03	-0.04	**0.71**	11.5%	1.9%	**72.4%**
Softgoal	0.10	**0.17**	0.05	17.3%	**21.2%**	10.1%

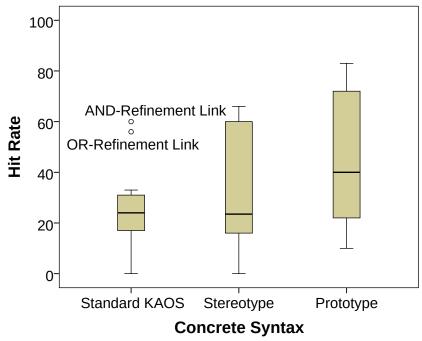
Descriptive statistics. Table 3 summarises the descriptive statistics for the collected metrics. Semantic transparency is defined as a scale from -1 to +1, and is measured by computing the semantic transparency coefficient of the symbols and the success of the participants matching the symbols from the 3 symbol sets to KAOS concepts. For each metric we present 3 rows in the table corresponding to the 3 symbol sets followed by the mean, standard deviation, skewness, kurtosis, and the p-values for the Kolmogorov-Smirnov and the Shapiro-Wilk normality tests. The *prototype* groups deviate from normality both concerning *semantic transparency* and *hit rate*, according to the Shapiro-Wilk test (p<.05). This is further illustrated through boxplots, presented in Fig. 4. Fig. 4a presents the semantic transparency coefficient, which is higher for the prototype symbol set. Fig. 4b shows the hit rate, which is also higher for the prototype symbol set.

Table 3: Descriptive statistics

Dependent Variable	Symbol Set	Mean	S.D.	Skew	Kurt	K-S	S-W
Semant. Transp. coefficient	standard	**.182**	.164	.845	1.256	.077	.114
	stereotype	.288	.409	.265	-1.440	.055	.057
	prototype	**.411**	.268	.386	-1.462	.110	.035
Hit rate	standard	**25.06**	15.105	.846	1.267	.091	.129
	stereotype	32.67	23.088	.269	-1.438	.051	.053
	prototype	**44.33**	25.294	.381	-1.465	.104	.034



(a) Semantic Transparency coefficient



(b) Hit Rate

Fig. 4: Results for the Semantic Transparency coefficient and Hit Rate

Hypotheses testing. We applied Levene's variance homogeneity test and found a statistically significant difference among the 3 distributions for semantic transparency coefficient (p=.007) and hit rate (p=.006). For comparing semantic transparency, we applied the Welch's t-test [24], which is robust when normality within groups and variance homogeneity among groups cannot be assumed. The semantic transparency for the 3 concrete syntaxes differs significantly according to Welch's t-test, $t(32.301)$=4.913, p=.014. The hit rate also differs significantly according to the Welch's t-test, $t(32.119)$=3.857, p=.032. This suggests that at least two of the concrete syntaxes differ significantly on their semantic transparency and hit rate. Post-hoc tests, using the Games-Howell post-hoc procedure, were conducted to determine which pairs of concrete syntaxes differed significantly. This test results are outlined in Table 4 and indicate that the semantic transparency of the prototype concrete syntax (M=.411, SD=.268) is significantly higher than the one of the standard concrete syntax (M=.182, SD=.164), with p=.12 and d=1.042. Concerning the hit rate, similar results are achieved, with the prototype concrete syntax (M=44.33, SD=25.294) significantly higher than the one of the standard concrete syntax (M=25.06, SD=15.105), with p=.026 and d=.938. The effect sizes for semantic transparency (d=1.042) and hit rate (d=.938) exceed Cohen's convention for a large effect (d=.80).

Our results suggest that the prototype concrete syntax is more semantically transparent than the standard concrete syntax. We found no statistically significant differences between the prototype and the stereotype concrete syntaxes, or between the stereotype and the standard concrete syntaxes. Also, the three concrete syntaxes are semantically transparent, even if in different degrees. The standard KAOS concrete syntax differs significantly from a semantically opaque concrete syntax (which would have a mean semantic transparency score around 0). We tested this with a one-sample t-test, $t(17)$=4.708, p<.001.

Table 4: Hypotheses testing for Semantic Transparency coefficient and Hit Rate

Hypotheses	**Formula**	**Mean Difference**	**Statistical Significance** *(p-value)*	**Practical meaningfulness** *(d)*
H_{1ST}	*stereotype > standard*	.106	.290	-
H_{2ST}	*prototype > standard*	.229	**.012***	**1.04****
H_{3ST}	*prototype > stereotype*	.123	.332	-
H_{4HR}	*stereotype > standard*	7.611	.480	-
H_{5HR}	*prototype > standard*	19.278	**.026***	**.938****
H_{6HR}	*prototype > stereotype*	11.667	.330	-

* The mean difference is statistically significant with $p < .05$

** Practical meaningfulness: $|d| > .8$, big effect size

8 Discussion

8.1 Evaluation of results

RQ1. Is the KAOS visual notation semantically opaque? The results do not allow us to conclude that the standard KAOS symbol set is semantically opaque. 67% of the participants of the semantic transparency experiment are from the WKML group. Some of them had contact with the KAOS language as part of a Software Engineering course, which might explain the relatively high semantic transparency coefficient values for the standard KAOS symbol set.

RQ2. Can participants with no knowledge in modelling languages design more semantically transparent symbols than participants with knowledge in modelling languages? The symbols produced by the WKML group are clearly influenced by the modelling languages they know, namely UML. The symbols produced by the WNKML group are less formal, more creative and different from each other. In the prototyping experiment (Study 3), the symbols drawn by the WNKML group had more votes than the ones drawn by the WKML group. In the semantic transparency experiment (Study 4), the prototype symbol set had significantly better results, which suggests the symbols drawn by the WNKML group were more easily identifiable. We conclude that the WNKML group produced symbols that represent better visual metaphors for KAOS concepts. Some participants had a background in Computer Science, while others did not. The former were significantly more able to produce a model with their proposed symbols than the latter but were less creative in their proposals.

RQ3. Which visual notation (*standard*, *stereotype*, or *prototype*) is more semantically transparent? The results show that there is a statistically significant difference between prototype and standard KAOS in terms of semantic transparency coefficient and success rate, with a considerable *effect size* for both metrics. We conclude that the prototype symbol set is more cognitively effective than the standard KAOS in terms of semantic transparency.

8.2 Implications to practice

The semantic transparency is only one of the 9 principles in the PoN. Improving a notation according to one particular principle does not necessarily lead to a more cognitively effective notation, as this change may have detrimental side effects with respect to other principles. For example, the ease of drawing the symbols is relevant for cognitive fitness, but is not considered here. Also, while the standard KAOS notation overloads some symbols, leading to a greater graphic economy, the prototype notation has more symbols, increasing the diagrammatic complexity. Although a symbol may be easily recognisable as mnemonic of a particular term, this may be a misrepresentation of a concept denoted by the same name, but with a significantly different semantics. For example, the symbols for obstacle and obstacle resolution were easily recognised by participants, but are not really related to the concept of obstacle in KAOS. Also, the symbols were evaluated in isolation, rather than in the context of requirements models. It may be the case that they form confusing diagrams, due to their conceptual diversity, as the metaphors were not chosen consistently from one symbol to the next.

8.3 Threats to validity

Conclusion validity. In the semantic transparency experiment, we used 18 candidate symbols for the stereotype and prototype concrete syntaxes, but only 12 for the KAOS standard concrete syntax, as it contains several symbols that overload different concepts [2]. This overloading introduces a bias for the smaller symbol set (standard KAOS) in terms of semantic transparency and hit rate. The probability of selecting the correct symbol by chance is higher for this set. This may have diminished the differences among the distributions of semantic transparency and hit rate. and may have hampered our ability to distinguish between the distributions of the standard and stereotype semantic transparency. The practical meaningfulness of the differences between the standard and prototype semantic transparency may be higher than measured in this experiment.

Internal validity. We targeted RE novices, using convenience sampling. Some participants had previously contacted with RE in an academic context, but all are surrogates for non-technical stakeholders and software developers who are not experienced in RE, thus controlling expertise bias. To mitigate sequencing effects, symbols were randomly ordered in the questionnaires for each participant.

External validity. We used novices to increase generalisability to the target population (stakeholders inexperienced with RE). As our participants are students from the same university, they share a common cultural background. Semantic transparency is often culture-specific, so their proposed and chosen concrete syntaxes were likely influenced by that background.

9 Conclusions and future work

We performed 3 quasi-experiments to support the evaluation of KAOS goal models semantic transparency and its improvement through the proposal of an alternative concrete syntax. We asked novices to draw candidate symbols for 18 KAOS goal model concepts. We created two alternative concrete syntaxes, based on these symbols: the stereotype and prototype symbol sets. Finally, we compared the semantic transparency of the 2 alternative concrete syntaxes and the standard KAOS, by asking a third group of novices to identify the symbol that better represents each KAOS goal model concept. The prototype's semantic transparency was significantly higher than the one in the standard KAOS concrete syntax (mean difference of .23). This suggests an opportunity for improving the communication between RE experts and other stakeholders using the prototype concrete syntax proposed in this paper. This result is in line with those obtained in similar studies for other modelling languages, as described in Section 2. Indeed, novices can be helpful in designing more recognisable symbols.

We plan to study other aspects of the PoN theory, such as complexity management, perceptual discriminability and cognitive fit. We also plan to assess if the prototype concrete syntax has drawbacks, in particular in model construction and model comprehension, since better symbol recognition may not necessary imply better model understanding. Moreover, since the symbols were selected independently from each other, they do not necessarily form a consistent set, in terms of the chosen visual metaphors.

Thus, further research is needed to study how an inconsistent set of symbols impacts the overall model understanding.

Acknowledgments

We thank NOVA LINCS UID/CEC/04516/2013 and FCT-MCTES SFRH/BD/108492/2015 for financial support.

References

1. A. van Lamsweerde, "Goal-oriented Requirements Engineering: A Guided Tour", International Symposium on Requirements Engineering, pp. 249–262, 2001.
2. D. L. Moody, "The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering", IEEE Transactions on Software Engineering, vol. 35, no. 6, pp. 756–779, 2009.
3. P. Caire, N. Genon, P. Heymans, D. L. Moody, "Visual notation design 2.0: Towards user comprehensible requirements engineering notations", RE'2013, pp. 115–124, 2013.
4. A. Dardenne, A. Van Lamsweerde, S. Fickas, "Goal-directed requirements acquisition", Science of Computer Programming, vol. 20, no. 1-2, pp. 3–50, 1993.
5. E. Yu, "Modelling strategic relationships for process reengineering", Social Modeling for Requirements Engineering, MIT Press, 2011.
6. A. Antón, W. McCracken, C. Potts, "Goal decomposition and scenario analysis in business process reengineering", CAiSE'94, pp. 94–104, 1994.
7. J. Castro, M. Kolp, J. Mylopoulos, "Towards requirements-driven information systems engineering: the TROPOS project", Information Systems, vol. 27, no. 6, pp. 365–389, 2002.

8. C. Rolland, C. Salinesi, "Modeling goals and reasoning with them", Engineering and Managing Software Requirements, 2005.
9. "GRL – Goal-oriented Requirement Language Kernel Description." `http://www.cs.toronto.edu/km/GRL/`, last access: March 2018.
10. A. van Lamsweerde, "Requirements engineering: From system goals to UML models to software", vol. 10. John Wiley & Sons, 2009.
11. A. van Lamsweerde, E. Letier, "From object orientation to goal orientation: A paradigm shift for requirements engineering", Radical Innovations of Software and Systems Engineering in the Future, pp. 325–340, Springer, 2004.
12. C. A. Gurr, "Effective diagrammatic communication: Syntactic, semantic and pragmatic issues", Journal of Visual Languages & Computing, vol. 10, no. 4, pp. 317–342, 1999.
13. D. Norman, "The design of everyday things: Revised and expanded edition", Basic Books (AZ), 2013.
14. N. Genon, D. Amyot, P. Heymans, "Analysing the cognitive effectiveness of the UCM visual notation", SAM'10, pp. 221–240, 2010.
15. N. Genon, P. Heymans, D. Amyot, "Analysing the cognitive effectiveness of the BPMN 2.0 visual notation", SLE'10, pp. 377–396, 2010
16. D. Granada, J. M. Vara, M. Brambilla, V. Bollati, E. Marcos, "Analysing the cognitive effectiveness of the WebML visual notation", Software & Systems Modeling, vol. 16, no. 1, pp. 195–227, 2017.
17. W. C. Howell, A. H. Fuchs, "Population stereotypy in code design", Organizational Behavior and Human Performance, vol. 3, no. 3, pp. 310–339, 1968.
18. S. Jones, "Stereotypy in pictograms of abstract concepts", Ergonomics, vol. 26, no. 6, pp. 605–611, 1983.
19. H. Zwaga, T. Boersema, "Evaluation of a set of graphic symbols", Applied Ergonomics, vol. 14, no. 1, pp. 43–54, 1983.
20. C. Howard, M. O'Boyle, V. Eastman, T. Andre, T. Motoyama, "The relative effectiveness of symbols and words to convey photocopier functions", Applied ergonomics, vol. 22, no. 4, pp. 218–224, 1991.
21. K. Arning, M. Ziefle, ""It's a bunch of shapes connected by lines": Evaluating the graphical notation system of business process modelling languages", 9th Int. Conf. on Work With Computer Systems, WWCS 2009. Beijing, China.
22. ISO, "Graphical symbols: test methods – part 1: methods for testing comprehensibility", Switzerland, 2014.
23. ISO, "Graphical symbols: test methods for judged comprehensibility and for comprehension. Geneva, 2001.
24. B. L. Welch, "The significance of the difference between two means when the population variances are unequal", Biometrika, vol. 29, pp. 350–362, 1938.
25. D. L. Moody, P. Heymans, R. Matulevičius, "Improving the effectiveness of visual representations in requirements engineering: An evaluation of i* visual syntax", RE'09, pp. 171-180, 2009.
26. R. Matulevičius, P. Heymans, "Visually effective goal models using KAOS", ER'07, pp. 265-275, 2007.
27. R. Matulevičius, P. Heymans, "Comparing Goal Modelling Languages: An Experiment", REFSQ'07, pp. 18–32, 2007.
28. F. Saleh, M. El-Attar, "A scientific evaluation of the misuse case diagrams visual syntax", Information & Software Technology 66, pp. 73-96, 2015.
29. S. Boone, M. Bernaert, B. Roelens, S. Mertens, and G. Poels. "Evaluating and improving the visualisation of CHOOSE, an enterprise architecture approach for SMEs", BRMP'14, pp. 87-102. Springer, 2014