Evaluating the Evaluators
An Analysis of Cognitive Effectiveness Improvement Efforts for Visual Notations

Dirk van der Linden and Irit Hadar
Department of Information Systems, University of Haifa, Haifa, Israel

Keywords: Requirements Engineering, User-Centered Software Engineering, Visual Notations, Modeling Languages.

Abstract: This position paper presents the preliminary findings of a systematic literature review of applications of the Physics of Notations: a recently dominant framework for assessing the cognitive effectiveness of visual notations. We present our research structure in detail and discuss some initial findings, such as the kinds of notations the PoN has been applied to, whether its usage is justified and to what degree users are involved in eliciting requirements for the notation before its application. We conclude by summarizing and briefly discussing further analysis to be done and valorization of such results as guidelines for better application.

1 INTRODUCTION

Conceptual modeling is a widely used technique to capture and reason about a particular domain of interest. The visual notation of a modeling language (i.e., its concrete syntax) is used to ensure that different stakeholders understand and agree on the same things. However, the design of visual notations for modeling languages is often based on intuition or committee consensus instead of empirical evidence. Some of the most widespread modeling languages used in practice like ER, UML and dataflow diagrams (Davies et al., 2006) suffer from such a lack of empirically grounded design rationale (cf. Moody and van Hillegersberg, 2009).

One of the main issues with visual notations developed in such ad hoc ways is a lack of focused attention on ensuring their *cognitive effectiveness*: the ease with which people can read and understand diagrams written in the newly developed or improved notation. Over the years, several frameworks have been proposed for evaluating, at least partially, this aspect. These frameworks provide notation designers with guidelines on how to better design visual notations. The frameworks range from relatively encompassing frameworks on multiple quality aspects such as the semiotics-based SEQUEL (Krogstie et al., 2006), to frameworks like Cognitive Dimensions, in particular its specialization for visual programming languages (Green and Petre, 1996) and Guidelines of Modeling (Schuette and Rotthowe, 1998). However, the intended focus of these frameworks differs, as well as their scope and practical use for analyzing visual notations instead of particular instantiations thereof (i.e., models written in them).

In 2009, Daniel Moody introduced a theory for cognitive effectiveness of visual notations, entitled the “Physics of Notations” hereafter referred to as the PoN (Moody, 2009). It is intended to deal with shortcomings introduced by other frameworks, in terms of evaluation scope and focus, and provide an evidence-based evaluation approach for designers to apply to new or existing visual notations. The adoption of the PoN framework by researchers is evident by the ever-growing number of analyses using it. Furthermore, a recent study has shown that while the number of research works using PoN is growing, the use of other, competing frameworks is simultaneously in decline (Granada et al., 2013).

With the growing significance of the PoN framework, ensuring its proper application becomes more important. Its prescriptive theory for designing cognitively effective visual notations consists of nine principles that are claimed to provide a scientific basis for the analysis and evaluation of visual notations. However, criticism has been expressed aimed towards the formulation of these principles and the difficulty of using the PoN in a replicable and systematic way (cf. Störrle and Fish, 2013; Gulden and Reijers, 2015; van der Linden, 2015; van der Linden and Hadar, 2015). In this paper, we will discuss our efforts invested so far on

Linden, D. and Hadar, I.
Evaluating the Evaluators - An Analysis of Cognitive Effectiveness Improvement Efforts for Visual Notations.
In *Proceedings of the 11th International Conference on Evaluation of Novel Software Approaches to Software Engineering (ENASE 2016)*, pages 222-227
ISBN: 978-989-758-189-2

investigating the use of the PoN. This investigation is based on a systematic literature review aimed at examining the thoroughness and scope of the applications of PoN, in order to identify systematic shortcomings in these applications, should they exist, and if so, how these shortcoming may be mitigated or resolved.

2 RESEARCH APPROACH

The goal of our study is to perform a systematic literature review (SLR) of work applying the PoN theory (Moody, 2009). We will follow the SLR guidelines proposed for applications in the Software Engineering (SE) field given by Kitchenham and Charters (2007). Specifically, the goal of this SLR is to investigate the applications of the PoN theory and analyze whether these applications have systematic shortcomings. Since the rigorous application of *scientific theory* to visual notation improvement in conceptual modeling is fairly new, it is important to endeavor that the work being done reaches its full potential. We thus focus on (1) articles applying the PoN theory to improve existing or new versions of notations, in terms of cognitive effectiveness; and (2) articles using the PoN theory as guiding principles during the creation of new modeling languages and notations. To the best of our knowledge, no SLR on the topic of applications of the PoN, nor similar frameworks in conceptual modeling, has been performed thus far.

2.1 Research Questions

The general research questions we address in our study are:

RQ1. What visual notations have been analyzed with the PoN theory?

RQ2. What justification for the use of the PoN theory is provided?

RQ3. To what degree do the analyses consider the requirements of their notation's users?

RQ4. How thorough are the performed analyses?

RQ5. Are there any (systematic) shortcomings in the applications of the PoN theory?

In order to answer RQ1, 2 and 4, we introduce below for each of the questions a number of sub-questions to be addressed when analyzing each of the primary studies included. The operational investigation of RQ3 and 5 will also be further elaborated on below.

With respect to RQ1, beyond identifying the specific notations analyzed, we wish to be able to differentiate between modeling tasks, which often call for different notations or use thereof, leading in some cases to the creation of multiple visual 'dialects'. This will be operationalized by identifying what the notation is used for; e.g., goals, processes, implementation or deployment. Furthermore, we wish to see how many new notations involve an *a priori* quality consideration. Thus, we distinguish between analyses of existing notations and analyses of new ones. Concretely, this results in the following sub-questions:

RQ1.1 What visual notation has been evaluated using the PoN theory?

RQ1.2 Is it an existing visual notation or a newly created one?

RQ1.3 What does the visual notation express (e.g., goal, process, rules)?

To answer RQ2 we wish to investigate the reasons reported for applying the PoN theory to the notation (i.e., whether it is called for), We operationalized this as follows:

RQ2.1 What reasons are given by the authors for analyzing the cognitive effectiveness of the given visual notation?

RQ2.2 What reason, if any, is given for the selection of the PoN theory over others?

RQ2.3 What alternative frameworks, if any, were considered?

For **RQ3** we focus on evaluating whether the analyses involved users in determining their requirements for the notation, i.e., if there is an explicit requirements phase involving actual or intended users of the visual notation before or during iterations of the notation design phase.

For **RQ4** we will investigate the thoroughness according to several criteria. First, since the PoN theory puts forth nine principles to analyze a notation by, we will investigate how many principles each analysis actually considered, keeping in mind that not all principles are equally relevant to all modeling contexts. This contextual evaluation is important so that the studied articles can be reasonably combined and compared (Khan et al., 2001). Second, we will analyze whether these principles have been considered in a systematic and replicable way. Finally, we will examine whether the concrete design suggestions stemming from the analysis were experimentally evaluated, and whether this evaluation involved actual (or intended) users of the notation. This leads to the following sub-questions:

RQ4.1 What is the scope of the analysis in terms of the PoN theory's nine principles?

RQ4.2 Was each included principle analyzed in a systematic, replicable way?

RQ4.3 Were the design suggestions evaluated as leading to measurable improvements for the cognitive effectiveness of the notation (e.g., higher reading speed, lower error frequency)?

RQ4.4 Did this evaluation include users of the notation? If no, how do the authors justify the results?

Finally, **RQ5** will be analyzed through tabulating the above findings, namely the ratio of analyses incorporating requirements elicitation and experimental evaluation, the average scope of the analysis and more. These meta-results will be examined to see if there are evident tendencies in the sample of selected papers; for example, a general absence of experimental evaluation or user involvement.

2.2 Search Process

It is important to ensure that a thorough search is done of appropriate databases and other potentially relevant sources (Greenhalgh, 2014). However, given our focus on analyses of existing or new notations via (partial) applications of the PoN theory, creating a search string that can effectively find them based on just title or abstract information is complicated (Brereton et al., 2007). Often many papers do not hint at the use of the PoN theory, or any analysis of the quality of the visual notation itself, instead using more vague and general terms in relation to the notation like its *quality* or *evaluation*. Thus, we decided to operationalize our search by searching for all papers citing the main publication of the theory (Moody, 2009). Operationalized, the search we used is thus:

`ALL PAPERS CITING` "The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering"

We used Google Scholar to search the articles to be included in the SLR, due to its demonstrated wide reach, which has been reported to return more primary sources than other comparable databases (Engström and Runeson, 2011), and has proven to be accurate in its recall in multiple domains (Gehanno et al., 2013). While Google Scholar has been more critically reviewed in the biomedical and medical domains as having lower recall than curated specialist databases (Bramer et al., 2013), these criticisms both assume the existence of a curated database specific to the field and queries yielding more than a thousand results, which does not come into play for our search. Furthermore, other work in software engineering has also successfully used Google Scholar as its exclusive means to extract cited-by information; see for example (Wohlin, 2014; Zhang and Babar, 2013; Zhang et al., 2011).

We incorporated manual curation based on a set of criteria to identify relevant articles (Zhang et al., 2011), so to verify that we did not miss published analyses that could be reasonably found. Potentially relevant articles were selected by the authors and vetted for relevance by each author based on the abstract and preliminary reading. This was done to ensure no conflicts of interpretation arose during the selection (cf. Da Silva et al., 2011). If any disagreements arose, we planned to ask impartial colleagues to give a tie-breaking opinion; however, no such disagreement has arisen so far.

2.3 Inclusion and Exclusion Criteria

Peer-reviewed articles and tech reports published by scientific institutions up to November 26th, 2015 that were found to have used the PoN, were included if they either:

- Reporting about applying the PoN theory, or a part thereof, to the evaluation of a visual notation.
- Discussing the applicability of the PoN theory, or a part thereof, to the notation at hand.

Articles with one or more of the following properties were excluded:

- No application or discussion of any part of the PoN framework.
- Papers published in a language other than English.
- Theses (bachelor, master or doctorate) unpublished in peer-reviewed sources.
- Overlapping versions of already included work. In this case the most complete paper was selected and used for the analysis.

2.4 Data Collection

The data we extracted from each paper included:

- Source and full reference
- Keywords
- Abstract
- The notation and its use (context of modeling)
- Scope of application: how many and which

principles were analyzed?

- Whether requirements were elicited, and if so, from whom?
- Whether an evaluation was done, and if so, with whom?
- Whether the paper provided a justification for the use of the PoN theory, and if so, what was it?
- Whether the paper discussed alternative theories to the PoN theory, and if so, which?

The first author extracted the data, which were checked by the second author. If there were any disagreements on the data, we resolved them via discussions, and had planned, if necessary, to involve impartial colleagues to give a tie-breaking opinion. So far no such disagreements occurred.

2.5 Data Analysis

The data were processed into a tabular overview to show:

- Year of publication
- Notation
- What is modeled by the notation (e.g., goal, process, implementation or deployment)
- Justification for using the PoN theory
- Inclusion of a priori requirements elicitation, operationalized as good, mediocre or bad to indicate: requirements elicited from target users, requirements elicited not from a different population than target users (e.g., students), and no requirements elicitation was done, respectively.
- Inclusion of experimental evaluation, operationalized as above.

It is important to note that we scored the occurrence of elicitation and evaluation steps, not taking into consideration the outcomes of these steps with respect to the evaluated studies' objective.

We then analyzed the scope of each application in terms of how many principles of the PoN theory were investigated. This was processed into a tabular overview and judged for each principle, operationalized as being well applied and reported, partially applied or reported, excluded, or unknown for indicating respectively: application of a principle with replicable description, application of a principle but no description of the means used, exclusion of the principle, and finally, those principles for which it cannot be verified whether the authors indeed applied it.

3 SEARCH RESULTS

According to the primary search criteria described above, the search resulted in a list of papers citing, at the time of writing, the Moody (2009) paper. This list included 502 articles. We then used per-year queries in Google Scholar, for each year of publication, in order to select papers for inclusion based on title, abstract, and preliminary reading. This led to an initial selection of 41 papers. Four of these papers selected on preliminary reading were excluded after analysis of the full paper, as no actual application of (any part of) the PoN theory was performed. This reduced the total number of selected papers down to 37, well in line with the expected range of retrieved primary studies for this kind of SLR (Kitchenham et al., 2009). Due to space constraints, the list of selected papers and extracted data is presented in an online Appendix at `www.dirkvanderlinden.eu/data`.

4 INITIAL FINDINGS

In this position paper we focus on a number of initial findings, which are potentially interesting in their own right and can be discussed in isolation.

4.1 Categories of Notations

We encoded the notations that the PoN was applied to, according to the following classification: (1) an existing notation, (2) a new notation, or (3) a new version of an existing notation. As can be seen in Fig. 1, there is a near balance between analyses of new and existing notations, with a far lesser ratio of analyses of new versions of existing notations.

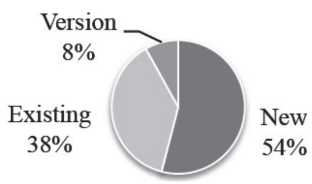
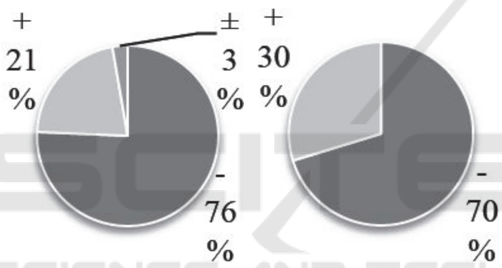


Figure 1: Ratio of new, existing, and versions of notations evaluated using the PoN.

This finding is important because it confirms that the PoN is not only used post hoc, but that notation designers are increasingly aware of its existence and potential benefits while designing a new visual notation. While the distinction between a new notation and a version of an existing one can be difficult, it makes sense to distinguish between the two as such versions often are mere dialectical changes of existing notations and share a significant part with their progenitor (e.g., the strongly related visual notations of goal modeling notations such as i*, GRL, KAOS).

4.2 Justification for using the PoN

Following from the previous finding, we examined to what degree authors justify using the PoN. That is, whether the choice for applying the PoN is made explicit and reasoned for, or whether it stays implicit. This also involves awareness of alternatives, such as other frameworks mentioned in Section 1. Fig. 2 shows the ratio of analyses justifying use of the PoN, and the ratio of analyses that considered alternative frameworks. The symbols +, +-, and – in Fig. 2a represent explicit reasoned justification, a brief and not reasoned justification, and no justification respectively, and similarly regarding the consideration of alternatives in Fig. 2b.



(a) Justifying **(b)** Considering alternatives

Figure 2: Ratio of analyses justifying their use of the PoN and considered alternatives.

Most analyses did not provide a justification for the use of the PoN, nor did they consider alternatives. Furthermore, when a justification was given, it often came down to repeating the justifications Moody himself had given for the creation of the PoN, rather than considerations originating from the authors. Analyzing the data, we found a large overlap between analyses that justify the use of the PoN and those that consider alternatives. This was indeed found to be the case, where all analyses that justified their use of the PoN also discussed at least one alternative framework, while two papers considered alternatives without finally giving a justification for their use of the PoN.

The high number of analyses that do not indicate the reason for using the PoN makes it difficult to investigate authors' reasons for doing so, as well as to what degree they are invested in proper application of the PoN, admittedly, a time and labor-intensive task.

4.3 Eliciting Requirements from Notation Users

A point of significant importance is whether authors using the PoN considered requirements posed by actual or intended users of the notation, in order to verify that the requirements set out by the PoN apply to their intended modeling task or users (van der Linden, 2015; van der Linden and Hadar, 2015). Fig. 3 presents an overview showing that very few analyses do so, with the majority never incorporating any explicit requirements elicitation or considerations.

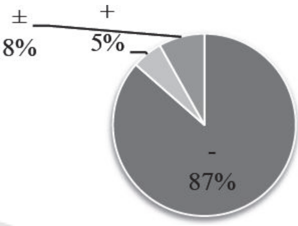


Figure 3: Ratio of applications of the PoN explicitly considering requirements of their users or modeling task.

While intuitively designing any artifact without considering its users' requirements seems problematic, from a pragmatic point of view an argument for avoiding the requirements elicitation step for this particular case can be made. Wiebring and Sandkuhl (2015) recently investigated requirements posed by users of business process modeling visual notations. They found that *"[...] a lot of these non-functional requirements closely resemble the principles constructed by Moody. For example, the demand for descriptive, graphic elements corresponds to the 'Principle of Semantic Transparency'."* Thus, while we do not wish to state in general that requirements gathering of users before the design of a visual notation is unnecessary, it might be the case that the PoN indeed pre-empts most (though not necessarily all) requirements that would be elicited.

5 CONCLUSION & OUTLOOK

This paper presented the research agenda and some preliminary findings of our SLR regarding the applications of the PoN. So far we have found that the PoN is applied more than assumed so far in literature (cf. Granada et al., 2013), having found thirty-seven applications for new, existing and versions of visual notations. This paper only discussed several initial findings, while the full results of our analysis cover a wider scope, dealing explicitly with evaluation and scope of the PoN's application. We intend to leverage on these findings toward better applications of the PoN. We will do so through the formulation of guidelines for aspects where the PoN applications can be improved.

REFERENCES

Bramer, W.M., Giustini, D., Kramer, B.M.R. and Anderson, P.F., 2013. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Systematic reviews,* 2, 1-9.

Brereton ,P., Kitchenham, B.A., Budgen, D., Turner, M. and Khalil, M., 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software,* 80, 571-583.

Da Silva, F.Q.B., Santos, A.L.M., Soares, S., Franca, A.C.C., Monteiro, C.V.F. and Maciel, F.F., 2011. Six years of systematic literature reviews in software engineering: An updated tertiary study. *Information and Software Technology,* 53, 899-913.

Davies, I., Green, P., Rosemann, M., Indulska, M. and Gallo, S., 2006. How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3), pp.358-380.

Engström, E. and Runeson, P., 2011. Software product line testing- a systematic mapping study. *Information and Software Technology*. 53, 2-13.

Gehanno, J.-F., Rollin, L., Darmoni, S., 2013. Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC medical informatics and decision making*, 13, 7.

Granada, D., Vara, J.M., Brambilla, M., Bollati, V. and Marcos, E., 2013. Analysing the cognitive effectiveness of the WebML visual notation. *Software & Systems Modeling*, pp.1-33.

Green, T.R.G. and Petre, M., 1996. Usability analysis of visual programming environments: a 'cognitive dimensions' framework. *Journal of Visual Languages & Computing*, *7*(2), pp.131-174.

Greenhalgh, T., 2014. *How to read a paper: The basics of evidence-based medicine*, John Wiley & Sons.

Gulden, J. and Reijers, H.A., 2015. Toward advanced visualization techniques for conceptual modeling. In *Proceedings of the CAiSE Forum 2015 Stockholm, Sweden, June 8-12*.

Khan, K.S., ter Riet, G., Glanville, J., Sowden, A.J. and Kleijnen, J., 2001. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews,* NHS Centre for Reviews and Dissemination.

Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J. and Linkman, S., 2009. Systematic literature views in software engineering – a systematic literature review. *Information and software technology,* 51, pp. 7-15.

Kitchenham, B. and Charters, S., 2007. Guidelines for performing systematic literature reviews in software engineering. *Tech Report EBSE-2007-01*. School of Computer Science and Mathematics, Keele University.

Krogstie, J., Sindre, G. and Jørgensen, H., 2006. Process models representing knowledge for action: a revised quality framework. *European Journal of Information Systems*, *15*(1), pp.91-102.

Moody, D. and van Hillegersberg, J., 2009. Evaluating the visual syntax of UML: An analysis of the cognitive effectiveness of the UML family of diagrams. In *Software Language Engineering* (pp. 16-34). Springer Berlin Heidelberg.

Moody, D.L., 2009. The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering. *Software Engineering, IEEE Transactions on*, 35(6), pp. 756-779.

Schuette, R. and Rotthowe, T., 1998. The guidelines of modeling–an approach to enhance the quality in information models. In *Conceptual Modeling–ER'98* (pp. 240-254). Springer Berlin Heidelberg.

Störrle, H. and Fish, A., 2013. Towards an Operationalization of the "Physics of Notations" for the Analysis of Visual Languages. In *Model-Driven Engineering Languages and Systems* (pp. 104-120). Springer Berlin Heidelberg.

van der Linden, D., 2015. An Argument for More User-Centric Analysis of Modeling Languages' Visual Notation Quality. In *Advanced Information Systems Engineering Workshops* (pp. 114-120). Springer International Publishing.

van der Linden, D. and Hadar, I., 2015. Cognitive Effectiveness of Conceptual Modeling Languages: Examining Professional Modelers. In *Proceedings of the 5th IEEE International Workshop on Empirical Requirements Engineering (EmpiRE)*. IEEE.

Wiebring, J. and Sandkuhl, J., 2015. Selecting the "Right" Notation for Business Process Modeling: Experiences from an Industrial Case. In *Perspectives in Business Informations Research* (pp. 129-144). Springer International Publishing.

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. Proceedings of the 18th Int. Conf. on Evaluation and Assessment in Software Engineering, ACM, 38.

Zhang, H. and Babar, M.A., 2013. Systematic reviews in software engineering: An empirical investigation. *Information and software technology*, 55 1341-1354.

Zhang, H., Babar, M.A. and Tell, P., 2011. Identifying relevant studies in software engineering. *Information and Software Technology,* 53, 625-637.