Van Der Linden, D., & Hadar, I. (2018). A Systematic Literature Review of Applications of the Physics of Notation. *IEEE Transactions on Software Engineering*. https://doi.org/10.1109/TSE.2018.2802910

A Systematic Literature Review of Applications of the Physics of Notations

Dirk van der Linden and Irit Hadar

Abstract—**INTRODUCTION:** The Physics of Notations (PoN) is a theory for the design of cognitively effective visual notations, emphasizing the need for design grounded in objective and verifiable rationale. Although increasingly applied, no systematic analysis of PoN applications has yet been performed to assess the theory's efficacy in practice. **OBJECTIVES:** Our primary objective was to assess the scope and verifiability of PoN applications. **METHOD:** We performed a systematic literature review (SLR) of peer-reviewed PoN applications. We analyzed what visual notations have been evaluated and designed using the PoN, for what reasons, to what degree applications consider requirements of their notation's users, and how verifiable these applications are. **RESULTS:** Seventy PoN applications were analyzed. We found major differences between applications evaluating existing notations and applications designing new notations. Particularly, in the case of new notations, we found that most applications adopted the PoN with little critical thought towards it, rarely considered its suitability for a particular context, and typically treated and discussed the PoN with few, if any, verifiable details and data. **CONCLUSION:** The results warrant consideration for those applying the PoN to do so carefully, and show the need for additional means to guide designers in systematically applying the PoN.

Index Terms—systematic literature review, physics of notations, visual notations, cognitive effectiveness, design rationale.

1 Introduction

Conceptual modeling is a technique used throughout most stages of information systems (IS) development to foster communication and shared understanding between stakeholders [1]. The requirements phase of software engineering (SE), for example, uses models as a major intermediary step in going from goals to actual software specifications [2]. Conceptual modeling languages and their products – the actual conceptual models – are extensively used to understand and communicate about a particular domain [3]. For these purposes, visual notations are frequently used rather than solely textual notations, as they can present information more concisely and precisely than a similar textual model [4], [5]. However, visual notations of modeling languages have frequently been designed by committees or ad hoc, without explicit regard for what empirical evidence would best suit a particular type of user and task. Some of the most widespread visual notations used in practice, such as the Unified Modeling Language (UML), have been claimed to be affected by this limitation [6].

Given the use of visual notations as tools for fostering shared understanding of, and agreement on, a given universe of discourse, it is important that they are easy to use and understand, and should therefore be designed to be cognitively effective [7]. Larkin and Simon [5] define cognitive effectiveness as "... the speed, ease, and accuracy with which a representation can be processed by the human mind." Many approaches exist that offer varying degrees of guidance in the form of guidelines and procedures to ensure the cognitive effectiveness of a visual notation. The application of a specific theory for designing cognitively effective visual notations, introduced by Moody [8], has grown increasingly widespread. Moody argues that the Physics of Notations (PoN) theory, offers a comprehensive work based on a synthesis of theories from, e.g., the psychology and cognitive science fields, to provide an evidence-based design theory that can be used in order to analyze the cognitive effectiveness of existing visual notations, or aid the design of new ones.

The main claim of the PoN theory is that grounding design choices in *design rationale* is vital for the development of any visual notation. In fact, the main article presenting the PoN [8] starts with a quote taken from [9], lamenting that "the reasons for choosing graphical conventions are generally shrouded in mystery." It then continues on to note that design rationale "is conspicuously absent in the design of SE visual notations," [8] and that, at the time of its writing, "SE visual notations are currently designed without explicit design rationale." [8] The solution, according to the PoN, is that justifications of visual notation design "should be based on scientific evidence rather than subjective criteria, as is currently the case." [8]

The goal of this review was to assess the state of the art of studies in the literature in which authors applied the PoN to visual notations, and determine whether the application of the PoN has led to its main goal: visual notations, the design of which is explicitly grounded in objective and verifiable design rationale. In an earlier publication [10], we presented preliminary results of the review with a subset of the applications described here, focusing on an initial descriptive synthesis. Here, we present the full results of a complete systematic literature review (SLR), incorporating a complete selection of studies.

Specifically, we investigated not only *what* those who

- *D. van der Linden is with the Department of Computer Science, University of Bristol, Bristol, UK. E-mail: dirk.vanderlinden@bristol.ac.uk*
- *I. Hadar is with the Department of Information Systems, University of Haifa, Haifa, Israel.*

Manuscript received December 3, 2016; revised August 9, 2017; revised November 27, 2017.

applied the PoN did, but also *how* they did it. To this end, this review was guided by four main research questions:

RQ1. Which visual notations have been analyzed using the PoN theory?

RQ2. What reasons do the researchers provide for using the PoN theory?

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

RQ4. How verifiable are the performed analyses?

The findings reveal that applications of the PoN are typically done with an uncritical attitude towards the theory, its suitability or need to be adapted to the notation at hand, and that design decisions informed by the PoN lack verifiability. We further discuss in detail the implications of the findings of this analysis and how they may steer future applications of the PoN. With this review, we hope to inspire researchers and practitioners, wishing to design or analyze visual notations according to the PoN, to make their efforts using the PoN as effective and reproducible for the conceptual modeling community as possible.

The remainder of this paper is structured as follows. In Section 2, we elaborate in more detail on the PoN theory itself, while Section 3 discusses related work critically assessing the PoN. Section 4 sets out our review approach by explicitly describing the research questions and the review protocol that followed from these questions. An overview of the data resulting from the analysis is given in Section 5, where each research question is individually answered. The meaning of these findings is discussed in Section 6. Finally, Section 7 summarizes the research contribution and concludes with recommendations for future use of the PoN.

2 Background: The Physics of Notations

The PoN theory has grown to become widely cited and referenced over the past few years, ostensibly welcomed by many notation designers for the guidance it provides.

This is evidenced not only by its high number of citations in a relatively small field (over 400 according to Scopus, and over 740 according to Google Scholar), which has steadily grown over the years, but also by an analysis showing that, as the number of works citing the PoN has grown, the number of citations to competing approaches, such as CD, has dropped [11].

Apart from the PoN's growing recognition in the conceptual modeling (language) literature, its influence has spread to other areas of research, such as conversation visualization [12], software interface design [13], model-driven development of statistical survey services [14], and analysis of musical composition [15]. Work has also been conducted to specialize it for use in analysis frameworks for specific domains, such as Enterprise Architecture [16]. This is not to say that the PoN has been universally adopted, as some authors have found its use unnecessary [17], while others found that applying parts of it may suffice [18] or even argue that their own proposed design principles, while in accordance with the PoN, are based on earlier sources and thus do not require its use (cf. [19], [20]).

The PoN was proposed as both a descriptive theory, to understand how visual notations communicate information, and a prescriptive theory, to improve notations' ability to communicate by ensuring that their design is cognitively effective. The prescriptive theory is the one that is most frequently referred to as "the PoN," consisting of nine principles: semiotic clarity, perceptual discriminability, semantic transparency, complexity management, cognitive integration, visual expressiveness, dual coding, graphic economy, and cognitive fit. Table 1 summarizes the principles.

TABLE 1
Overview of PoN principles, adapted from [8].

#	Principle name	Summary
1	Semiotic clarity	There should be a 1:1 correspondence between semantic constructs and graphical symbols
2	Perceptual discriminability	Different symbols should be clearly distinguishable from each other
3	Semantic transparency	Use visual representations having an appearance that suggests their meaning
4	Complexity management	Include explicit mechanisms for dealing with complexity
5	Cognitive integration	Include explicit mechanisms to support integration of information from different diagrams
6	Visual expressiveness	Use the full range and capacities of visual variables
7	Dual coding	Use text to complement graphics
8	Graphic economy	The number of different graphical symbols should be cognitively manageable
9	Cognitive fit	Use different visual dialects for different tasks and audiences

Each of these principles focuses on a particular aspect that contributes to the cognitive effectiveness of a visual notation, providing theoretical definitions and procedures to verify or implement the aspect for a given notation. However, the detail and degree to which procedures are given to explain the practical implementation of each principle vary. Some are fairly straightforward, some require understanding of their related cognitive theories, and some require user involvement in order to both operationalize and implement them [21]. Moreover, the implementation of one principle may affect, positively or negatively, the implementation of another principle. An overview of these interactions as established by Moody is shown in Fig. 1.

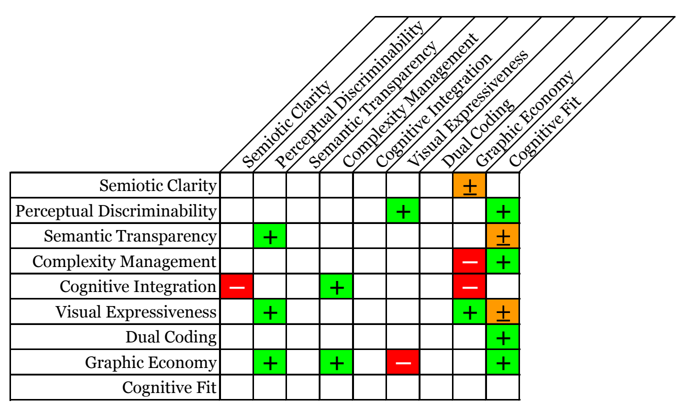


Fig. 1. Trade-offs between PoN principles, from [8].

Some principles require only the elements of the visual notation and its related semantics to be applied. For semiotic clarity, one has to compute some formulae for symbol redundancy, overload, excess, and deficit. Together, they indicate whether it holds that each semantic construct has a unique graphical symbol, used only for that construct. This requires access to the specification of the language for the constructs and the visual notation, but otherwise requires little additional effort.

Sometimes the application of a principle that requires theory is straightforward. To compute *graphic economy*, we need to check that the total number of graphical symbols respects some threshold of being cognitively manageable. Practically speaking, this is done by comparing the total number to Miller's law of 7 ± 2 symbols. If the total number of symbols is ≤ 9, one could already argue the principle holds. Of course, if the visual notation contains different diagram types, in line with *cognitive integration*, the calculation becomes slightly more involved, as one would need to calculate the total number of graphical symbols that can potentially occur in each diagram type.

Not all applications of theory are as simple as they seem, however. Take *dual coding* for instance. Using text to complement graphics sounds simple enough; but what additional design choices should be made to ensure that the text is also readable and effective? Where should a textual label be located? Should it have a color that maximizes the contrast with the element on which it is placed? Should its size be a consideration, or even considered a static property if models can be zoomed in and out? In terms of how involved the implementation and design rationale are, there is thus a long continuum between naïve and more thoughtful applications.

Finally, consider perhaps one of the most deceivingly simple principles: *semantic transparency*. The visual notation should use graphical symbols that suggest their meaning. We should use icons that represent the real-world element, like using a stick figure for human actors, a lock element for secured objects, and so on. How does one implement this principle? User involvement becomes necessary both to elicit potential meaningful symbols and to evaluate whether their suggested meaning holds for the intended user audience of the visual notation. Given the PoN's insistence that design rationale should be based on scientific evidence rather than subjective criteria, implementing these kinds of principles requires extensive empirical work.

Given all these considerations, one could thus ask whether the PoN indeed contains a prescriptive theory or whether it is a collection of guidelines and best practices stemming from relevant theories, however open for interpretation by its users. It is likely that one application of the PoN would not be informed by the exact same interpretation of its principles and their scope as another.

3 Related Work

While the attention paid to the PoN has grown, so has the research critically analyzing its suitability and capability as a theory for determining the cognitive effectiveness of visual notations. The PoN has been criticized for being overly focused on visualizations that deal with individual entities and their relationships, and for not offering sufficient support for other paradigms, such as pattern-based visualizations [22]. da Silva Teixeira et al. [23] argued that the PoN itself does not offer a systematic process for applying the theory and that it lacks guidance in terms of explaining exactly when to apply which principle.

Störrle and Fish [24] criticized the formulation of the PoN's principles for being "neither precise nor comprehensive enough to be applied in an objective way to analyze practical visual software engineering notations." They attempted to alleviate this issue by providing formalizations to verify whether a given principle holds, focusing on two principles: semiotic clarity and perceptual discriminability. However, even for their proposed formalizations they encountered the challenge of needing to make choices concerning particular variables and threshold values, for which they acknowledged "not yet having the empirical evidence to support our assumptions." [24] Some effort in this direction has been made by Stark et al. [25], who provided theoretically-grounded color palettes, aiding the application of the perceptual discriminability principle. CEViNEdit, a recently proposed editor for the creation of domain-specific languages that claims to take the PoN into account [26], was similarly limited, focusing on three principles: semiotic clarity, visual expressiveness, and graphic economy.

In relation to these limitations, van der Linden et al. [21] investigated the degree to which the principles of the PoN lend themselves to operationalization efforts. They concluded that it is impractical, if not downright impossible, to capture several of the principles in finite formalizations. One of the reasons was that user involvement is required for both the operationalization of some principles and verifying whether these principles hold. This provides further grounds for a critical investigation of applications of the PoN theory, such as the reported finding that applications frequently lack the involvement of the intended users of the visual notation [27]. To support designers of visual notations, van der Linden et al. further proposed a framework [28] which explicitly acknowledges the claimed impracticality of operationalizing several principles, instead guiding designers to apply the PoN while explicitly documenting design rationale and noting supporting evidence and the strength thereof.

While it can thus be said that research has paid ample attention to ensuring the quality of the PoN itself, the same cannot be said for the way it is used. Little to no work exists that has looked at applications of the PoN and the verifiability of their claimed design rationales. Yet, the consequences of some of the criticism directed at the PoN, such as the lack of user involvement or the lack of precision in the formulation of the principles, would be more effectively investigated and reasoned about in the actual context of use.

Given the increasing attention paid to the PoN, this is of importance, because the theory should be evaluated not only on its own merit, but also in its actual context of use. Since the rigorous application of scientific theory to visual notation design in conceptual modeling is fairly new, it is important to endeavor that such design reaches its full potential.

4 Review Approach

4.1 Background

The goal of our study was to gain insights into how the PoN is applied. To do so, we performed an SLR of work in which the PoN theory was applied, concretely assessing its application by investigating the scope of applications of the PoN, and the way each principle was applied. We follow the SLR guidelines proposed for the field of Software Engineering (SE) given by Kitchenham and Charters [29]. We focus on (1) studies that applied the PoN theory in order to improve existing notations in terms of cognitive effectiveness, and (2) studies that used the PoN theory during the creation of new modeling languages and notations. As far as we know, no SLR on the topic of applications of the PoN or similar approaches in conceptual modeling has yet been performed.

Because we want to have an effect on the practice of visual notation improvement, it is important to ensure that the insights provided by this SLR are meaningful and important to practitioners as well as to researchers [29]. We aim to ensure this by identifying the gaps in the application of the PoN theory to existing or newly proposed visual notations, on the basis of which we formulate recommendations for using the theory more critically, emphasizing the requirement of involving the actual or intended users of the notation. These strategies are constructed with the objective of leading to notations demonstrating higher cognitive effectiveness for their actual users.

Kitchenham et al. [30] found that a relatively large number of studies were focused on questions that are relevant mostly to research. While our sample focuses on peer-reviewed studies from scientific sources, the questions we phrase are directly relevant to practice, as we are concerned with the involvement of practitioners in the design of the artifacts they use. Kitchenham et al. [30] further found that SLRs aimed at evaluating technology (e.g., models, methods, tools) often incorporate a lower number of primary studies (ranging from 6 to 59) than those investigating research trends (ranging from 63 to 1485). Given our focus on applications of the PoN to visual notations, we thus would expect the number of primary studies to fall within the range found for evaluating technology via SLRs.

The most important issue in the current use of SLRs that Kitchenham et al. [31] identified is the lack of quality assessment in the primary studies and thus the reliability of their findings. Given this need to assess the quality of the primary studies that applied the PoN theory, we explicitly formulated research questions that assess the quality of the work accomplished in these studies.

4.2 Research Questions

To paint a complete overview of the landscape of applications of the PoN, we investigated several aspects, each addressed by their respective research questions. We addressed the following research questions to be answered by this SLR.

RQ1. Which visual notations have been analyzed using the PoN theory?

Some authors who applied or discussed the PoN mentioned some examples of other applications, but none gave a complete overview that identifies the visual notations to which the PoN was applied and what this landscape looks like. In particular, we are interested in finding how many applications of the PoN were focused on analyzing existing notations and how many authors used the PoN as guidance during the design process of new notations. Given the intended impact of the PoN as a Design Theory that can be used to improve the design of new notations, it is of interest to see the extent of the impact it has generated in that direction. Furthermore, we were interested in identifying in what aspects (e.g., goals processes, rules, architecture) the visual notations were used, as this may lead to interesting classifications of other results later, for example, whether a particular aspect, such as process modeling, typically focuses more on some principles than do other aspects. Specifically, this results in the following sub-questions:

(a) Is the notation an existing or a newly created one?
(b) For what aspects (e.g., goal, process, rules) is the notation used?

RQ2. What reasons do the researchers provide for using the PoN theory?

With regard to RQ2, our aim was to identify the reason for which authors chose the PoN theory rather than other existing approaches. To answer this question, we note what reasons, if any, are given for choosing the PoN and whether any alternative approaches (e.g., CD, GoM, SEQUAL) were considered. This could give some insight into what aspects of the PoN (i.e., its theoretical nature and concrete focus on visual notations) have led to it receiving more attention than other approaches. Specifically, this results in the following sub-questions:

(a) What alternative approaches, if any, were considered?
(b) What reasons, if any, are given for the selection of the PoN theory over others?

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

We wanted to understand whether authors who design a new visual notation or analyze the cognitive effectiveness of an existing one consider user requirements. While the PoN provides a strong theoretical basis for analyzing cognitive effectiveness, differences in what users in a particular domain perceive as most important could affect the application of the PoN by, e.g., providing more context-specific information for the operationalization or prioritization of principles. We investigated whether authors take into account empirical data elicited from intended users of the notation via the following sub-questions:

(a) To what extent do analyses involve users in determining their requirements for the notation?
(b) For each principle, to what extent are users involved in operationalizing the principle?
(c) To what extent are trade-offs between principles discussed and determined with the involvement of users?
(d) Were the design decisions regarding visual elements evaluated such that the measurable improvements in the cognitive effectiveness of the notation, as compared to its alternative design, were demonstrated?
(e) Did this evaluation of the designed notation include the intended users of the notation? If not, how do the authors justify their evaluation procedure?

RQ4. How verifiable are the performed analyses?

We investigated the verifiability of each PoN application by focusing on what was in fact analyzed and how it was reported. First, because the PoN theory defines its nine principles as the independent variables that affect a notation's cognitive effectiveness, we investigated the number of principles each application in fact considers, bearing 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 [32]. Second, we investigated whether the application of the selected principles was reported such that others can verify and understand exactly what was done. This leads to the following sub-questions:

(a) What is the scope of the analysis in terms of the PoN theory's nine principles?
(b) Is the analysis of each principle verifiable?

4.3 Review Protocol

4.3.1 Data sources and search strategy

To the extent of our knowledge, no SLRs on applications of the PoN exist. Some applications of the PoN mention other applications in a related work section. This typically involves the same small number of PoN applications to major modeling languages, such as WebML, BPMN, and UML (although the typically cited paper for the analysis of UML predates the PoN and uses a non-definitive version of the PoN still under review at the time - the very reason why it was not included in this review). Thus, there is no prior published work to use as a starting point for this SLR.

Instead, we started our search through digital libraries. Given the relatively recent publication date of the PoN (2009), we did not foresee limiting ourselves to digital libraries as a handicap. Creating a search string that can effectively find applications of the PoN based only on the title or abstract information is complicated. Frequently, authors do not hint at the use of the PoN theory, or any analysis of the quality of the visual notation itself, in the abstract, but rather use more vague and general terms in relation to the notation, such as its quality or evaluation.

Thus, we decided to first construct and trial our search query, using Google Scholar, by searching for papers that mention either the *name* of the theory or its *dependent variable*. This resulted in the following query:

"Physics of Notations" OR "cognitive effectiveness"

This resulted in a long list of papers, many of which were not related to modeling languages or to visual notations at all. In several iterations, we decided to add relevant terms to the right part of the disjunction to enforce a correct context. These terms were derived from PoN applications found mentioned in related work sections. This led to the final query:

("Physics of Notations" OR "cognitive effectiveness") AND ("modeling" OR "language" OR "notation" OR "visual" OR "diagram")))

We checked for papers from 2009 onward to ensure comparability, as some earlier papers, including one authored by Moody himself (analyzing UML), predate the definitive publication of the PoN in IEEE TSE. The PoN underwent several changes over its earlier versions, including different names of its principles (e.g., perceptual immediacy vs. semantic transparency). These cannot be trivially compared to other papers, as the version of the PoN to which the authors had access would have been different.

We performed our search query in seven databases: the ACM DL (12 results), IEEE Xplore (10 results), SpringerLink (185 results), ScienceDirect (45 results), the AIS Electronic Library (59 results), Web of Science (12 results), and Scopus (622 results). Then, we ran a forward snowball on the PoN, extracting a list of papers citing [8] from Web of Science, Scopus, and the ACM Digital Library (see Fig. 2 for an overview). Scopus was the most complete in this regard, with no other database having any additional cited-by data to add. No new results were found in this step.

Additionally, we used Google Scholar to verify that we did not miss any articles not indexed in the selected databases. Because of its wide reach, Google Scholar can be a valuable source for such efforts, having been noted to help in the retrieval of even the most obscure information [33]. This step added two journal articles not found by any of the previously used databases because of an interesting combination of circumstances. The titles and abstract were written such that our search query did not pick up on them because of a lack of detail. Normally, they would have been found regardless of this fact through an all-fields search, because they reference the Physics of Notations. However, these two articles were published online in Software & Systems Modeling and had not yet, at the time of the search, been assigned to a specific volume. This caused Scopus and other databases to treat them as "in press" and not index reference data, making it impossible to find them with the "Physics of Notations" part of the query. Using Google Scholar's wider (albeit far less curated) reach, we managed to include these articles, regardless of the described limitation.

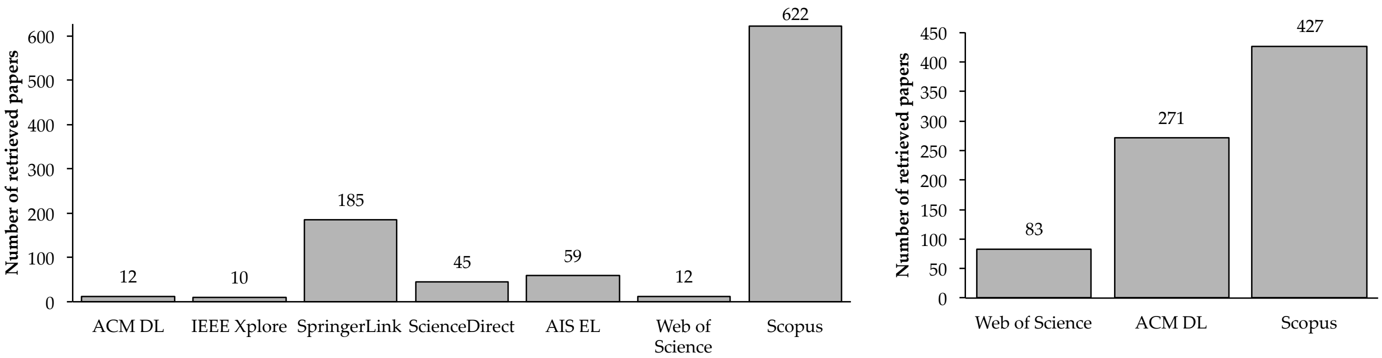
Search strategy publication bias

While we included only published, peer-reviewed work, we searched through all typical levels of publications from journal to conference to workshop. This should have yielded a representative selection of studies of varying level of maturity (and perhaps, quality). Workshops and conferences in the SE and IS domains typically include tracks to discuss work-in-progress focused on providing authors with feedback. Thus, this should feasibly have absolved to some degree the inherent publication bias by reducing the threshold for publication from solely journal articles.

Search strategy documentation

We documented each search result by exporting generated result lists from the used databases. These were saved as comma-separated or plaintext files and are available in an online appendix.[1]

1. See: https://is-web.hevra.haifa.ac.il/files/vanderlinden/vanderlinden_pon_slr_search_results.zip.



(a) Retrieved search query results **(b) Retrieved forward snowball results**

Fig. 2. Returned papers from each database for (a) the search query, and (b) the forward snowball of papers citing [8].

4.3.2 Study selection

Study selection criteria

We selected peer-reviewed studies published from 2009 up to May 15th, 2017 if they satisfied *all* the following inclusion criteria:

1) They described the development and/or evaluation of one or more visual notation(s), and
2) They described an unmodified application of (a part of) the PoN for the development and/or evaluation of that visual notation.

In addition to the step-wise inclusion criteria, we excluded articles if they

3) had overlapping versions of already included work. In this case, the paper with the most complete description of the application of the PoN (thus not necessarily the most recent chronological paper) was selected and used for the analysis.

Study selection process

Fig. 3 gives an overview of the search and selection process, with the number of remaining studies after each step. Ultimately, the selection process resulted in a set of 70 papers, slightly above the range of retrieved primary studies that could be expected according to Kitchenham et al. [31]. A full list of the selected papers is given in Table 2.

The inclusion criteria were treated step-wise. That is, first we checked whether the paper described the development and/or evaluation of one or more visual notation(s), and if so, only then did we check whether it described the application of (a part of) the PoN.

Second, in line with Kitchenham's guidelines, we *liberally* applied the inclusion criterion of describing the development and/or evaluation of one or more visual notations. Brereton et al. noted that "The standard of IT and software engineering abstracts is too poor to rely on when selecting primary studies." [34] Kitchenham thus suggested that "Initially, selection criteria should be interpreted liberally, so that unless studies identified by the electronic and hand searches can be clearly excluded based on titles and abstracts, full copies should be obtained." [29] Practically, this meant allowing for the inclusion of articles found by the query where the abstract noted developing, e.g., a tool, framework, or method, where it could reasonably be expected that such an effort would include a notation, even if not explicitly mentioned in the abstract.

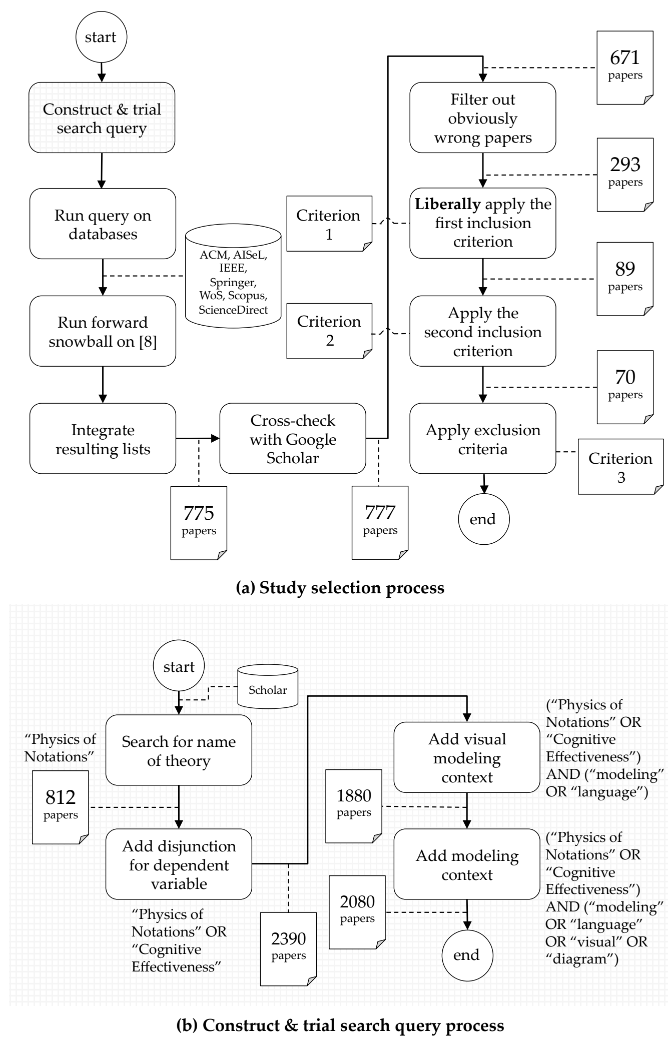


Fig. 3. Flowchart of (a) the study selection process and (b) the construction and trial of the search query. The search query was constructed using Google Scholar, explaining the inflated counts of returned papers relative to used data sources in the actual selection process.

Third, we checked whether the studies applied (parts of) the PoN unmodified, taking the theory as published. We made this refinement after coming across several studies that applied a synthesis of the PoN and other approaches. This made it infeasible to separate out what parts of the visual notation design was informed by the PoN's principles. Studies that applied (a part of) the PoN and complementarily applied other approaches were included, as it was feasible to separate out what was done according to which approach.

The inclusion criteria required some interpretation. For the papers retrieved from the first three databases (ACM, IEEE, Web of Science), both authors independently assessed the selection criteria and then discussed them, finding no disagreements. Because of this high level of agreement, we decided to continue with the first author assessing and documenting the remaining retrieved results. At the conclusion of this stage, we selected a random subset of 30 papers for the second author to re-assess the inclusion decisions and compare them with their own decision. No disagreements were found, allowing us to move on to the exclusion criteria.

In line with Kitchenham's guidelines, we did not a priori exclude papers in languages other than English. Several articles were found in French, German, and Spanish. These were read by the first author who speaks French and German, and corroborated with native speakers. For the articles in Spanish, we consulted a native speaker to verify whether the inclusion criteria held. In all cases, even if these papers satisfied all inclusion criteria, they were excluded later because a (more complete) English version detailing the same work had also been published.

Study selection documentation

We documented each phase of the study selection process. A comma-separated file for each searched database is available in an online appendix[2], explicitly capturing the representative quotes used to ground the decision for each study to (dis)satisfy the criteria. Fig. 4 gives an overview of the number of papers included per year, publisher, and venue.

4.3.3 Study quality assessment

Whether it is necessary to explicitly assess the quality of primary studies is dependent on the type of SLR that is undertaken [34]. Because the objective of this review is to assess PoN applications, the actual data extraction and synthesis focus on establishing the quality of the studies. This was done by determining whether a priori requirements elicitation for the visual notation was done (i.e., appropriateness), design choices were evaluated (i.e. validity), and the implementation of design according to the PoN principles is verifiable (i.e., validity).

Furthermore, an important aspect of our quality assessment is establishing whether claims are verifiable. Given the PoN's purpose as a design theory, providing rationale and argumentative proof for design choices is of the utmost importance. For this reason, we coded the level of verifiability of each PoN principle, establishing an additional quality metric, in particular, by explicitly capturing when claims are unverifiable. Making this distinction is also seen as important by Kitchenham, who stresses the need to not assume that "because something wasn't reported, it wasn't done." [29] While it is true that one cannot verify that an unverifiable PoN principle was not implemented, for all intents and purposes one cannot assume it either. Thus, such applications offer little value for the later selection of a visual notation that satisfies some given set of principles.

Finally, we chose not to incorporate further *a priori* quality cohorts based on studies' metadata. For example, we could naively divide papers into cohorts according to their publication venue, with journals being ranked above conferences, and conferences above workshops. Journals could then be ranked according to, e.g., impact factor, and conferences and workshops according to published rankings. However, regardless of whether such quantifications are actually valid indicators of quality, there is the additional difficulty of assessing the different quality level between cohorts: how does a highly-ranked conference study relate to a study in a very low impact factor journal?

4.3.4 Data extraction

Design of data extraction forms

We used a spreadsheet to capture the extracted data, including:

- Title
- Year 2009, . . . , 2017
- Venue journal, conference, workshop
- Source
- Publisher

2. See: https://is-web.hevra.haifa.ac.il/files/vanderlinden/vanderlinden_pon_slr_inclusion_decisions.zip.

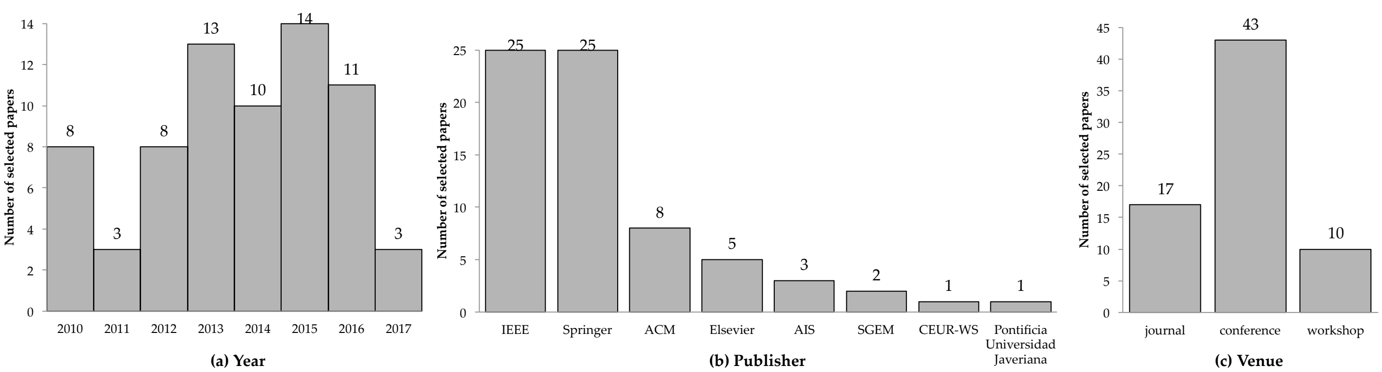


Fig. 4. Overview of the number of included studies (a) per year, (b) by publisher, and (c) by venue

- Authors
- Name of the visual notation designed or evaluated
- Context of use of the visual notation
- Novelty status {existing, version, new}
- Were requirements elicited, and from whom {intended users, others}?
- Was an evaluation of the design done, and with whom {intended users, others}?
- What justification was given for the use of the PoN, if any?
- What alternatives to the PoN were discussed, if any?
- What was the scope of the PoN application: how many, and which principles are applied?

The first author extracted the data, which were checked by the second author. As no disagreements arose during the data extraction stage, no further procedure for resolving issues was needed or applied.

Data extraction from external files

Authors do not always include all details in the actual publication itself. We allowed for the extraction of data from external files, such as technical reports and online datasets, on the condition that the selected study explicitly provided directions to external files in the publication. Examples of such applications are Papers 6, 7, and 45.

4.3.5 Data synthesis

Descriptive synthesis

The data were processed into a tabular overview (see Table 3) to show:

1) Notation
2) Focus
3) Status
4) Elicitation of requirements
5) Evaluation of design
6) Justification of PoN use
7) Investigated alternatives

It is important to note that we scored the occurrence of elicitation and evaluation steps and not whether the outcomes of these steps in the primary studies were positive or negative with respect to the evaluated studies' objective. For the elicitation of requirements and evaluation of design, we used three possible codes: + for doing so from the intended users of the notation, ± for doing so from participants who were not necessarily the intended users (e.g., students), and - for not doing so. Justification of PoN use was simply coded as a + for any explicitly given justification, - for none. Investigated alternatives were coded as either the abbreviations of the mentioned approach, or - for no mentioned alternatives.

Qualitative synthesis

To assess the scope of the PoN analysis, an additional tabular overview (see Table 4) was created. Here, we classified the status of each PoN principle with four codes: + for good, ± for somewhat, - for excluded, and ? for unverifiable. This operationalization respectively refers to the case: + if a principle was claimed to be used, and a verifiable rationale was present; ± if a principle was claimed to be used, but no rationale was presented, or described such that subjective interpretation would be necessary (e.g., because of no explicit mapping to either the direct principle or terminology used by Moody); - if a principle was explicitly claimed to be not applied, or irrelevant for the notation; and finally, ? if the application of a principle cannot be verified. For our intents and purposes, a classification of ? is worse than - , because no rationale can be distinguished and it cannot be ruled out that a particular principle was meant to be applied. Instead, the effect of the principle on the visual notation becomes unverifiable, as discussed in Frankfurt's treatise on the epistemic condition of propositions of which the truth status cannot be verified [35].

Both authors individually coded the principle scope according to this scheme. Then, we compared the classifications and resolved any disagreements. We calculated a kappa co-efficient of 0.934 with an SE of 0.012, giving a "very good" strength of agreement in the 95% confidence interval for the authors' respective classifications.

Sensitivity analysis

Sensitivity analysis is somewhat more subjective in our case given the use of descriptive and qualitative synthesis. However, as part of this review's objective we needed to investigate differences between cohorts of papers, as discussed in Sec. 4.3.3. We found that there is a difference in terms of principle scope between journal articles and proceedings (conference and workshop) papers. In journal papers, only 12% of the principles are unverifiable, whereas in conference and workshop papers the percentage is markedly higher (32% and 37%, respectively). This difference is taken into account in the interpretation of the results, where we take care not to generalize to all papers.

Additionally, we also checked sensitivity for the type of study, i.e., whether it focused on analyzing an existing notation or on designing a new (version of a) notation. A major difference exists between the number of unverifiable principles in applications to new/versions of notations (38%) and applications to existing notations (7%). This is likely an indication of the objective of PoN applications: the analysis of existing notations is aimed to improve something, whereas in applications to newly designed ones, the PoN may simply be treated as an afterthought.

5 Results of the Review

The data resulting from our analysis are given in the Appendix in Tables 3 and and 4. Table 4 shows the classification of all aspects up to, but not including, the question of which principles were addressed by a paper. It gives an overview of whether each principle was addressed (and to what degree) by a paper. The following subsections answer each research question and its sub-questions in detail.

5.1 Which visual notations have been analyzed with the PoN?

The visual notations investigated with the PoN are shown in Table 3. Our results show that the number of papers that discuss an application of the PoN as compared to the total number of papers citing it is not very high: 16%

(based on Scopus' cited-by number of 427). In many cases, such non-application citations are general references in related work or used as a general reference to indicate that visual notation is important or that some particular aspect or visual variable (i.e., color use) is important. The total number of visual notations analyzed in the selected papers is slightly higher than the number of selected papers, because Paper 3 reports analyses of three distinct visual notations for similar purposes. In some papers reporting studies where a new visual notation was created, the authors did not explicitly name or otherwise identify their notation, e.g., referring to it as "a visual notation for [topic]." Nonetheless, we consider them equally to other, named, notations.

RQ1a: Is the notation an existing or a newly created one?
From Table 4, we can derive the relative number of newly created, existing, or versions of existing visual notations that were reported in the reviewed papers. These ratios are presented in Fig. 5(c). There seems to be a balance between use of the PoN for entirely new visual notations and for previously existing ones, which includes both analyses of existing notations and applications of the PoN to design versions or "forks" of existing notations, to create an extended version of a notation already in existence, such as in Paper 39, reporting on "extended Compliance Rule Graph."

RQ1b: For what aspects (e.g., goal, process, rules) is the notation used?
There is a large plurality of notations judging from the purposes for which they are used, ranging from well-known modeling foci, such as business processes, goals, and requirements, to more specialized foci, such as test environments, situational dependencies, and so on. While there is a certain overlap between the stated focus of visual notations reported in the selected papers, such as multiple notations dealing with (different aspects of) business processes, it is difficult to categorize them into a small set of modeling foci without risking arbitrary categorization choices

5.2 What reasons do the researchers provide for using the PoN theory?

RQ2a: What alternative approaches, if any, were considered?
While from the above results it can already be inferred that only a small number of alternative approaches were considered, Table 3 lists several such approaches, which are visualized in Fig. 5(b). The few papers in which alternatives were considered typically mentioned the CD approach, following Moody's treatment and description of CD as the prime alternative to the PoN. Other approaches mentioned are SEQUAL [36] and Guidelines of Modeling [37]. Some papers that focused on specific topics mentioned more specialized approaches, such as the 7PMG [38].

RQ2b: What reasons, if any, are given for the selection of the PoN theory rather than others?
We investigated this question by assessing first which papers provide a justification for applying the PoN, and second which papers mention or discuss alternative approaches. Fig. 5(a) shows the ratio of papers in which the authors justify their use of the PoN theory with some explicit arguments, coded as +, justify it without explicit arguments, coded as ±, and provide no justification coded as -. Fig. 5(b) shows the ratio of papers where the authors considered any alternative approaches, coded as +, to those where the authors did not, coded as -. While the populations of Fig. 5(a) and (b) do not completely overlap (see Table 4 for more details), there seems to be a connection between those where the authors give explicit justification for selecting the PoN and those where they considered alternatives. Most authors who considered alternative approaches also gave a justification for having chosen the PoN.

Only 14% of the papers contained explicit detailed reasons for analyzing the cognitive effectiveness of their notation, short of paraphrasing Moody's clear thesis on the need for cognitively effective notations. This is particularly the case for newly created visual notations. For example, in Paper 16 the authors state the following as the reason for the analysis "for the visual language proposed here, the PoN principles are applied because of their scientific and theoretical validity." Others justify their use primarily by its (perceived) widespread use, e.g.: "we chose PoNT[heory] as it is the state of the art SE and RE notation evaluation frameworks widely used with other notations." [39]

An example of a well justified use of the PoN can be found, for example, in Paper 6, where the authors justified the use of the PoN by discussing the limitations of alternative approaches, and presented an argument why the PoN was their best option. First, CD and its relevant

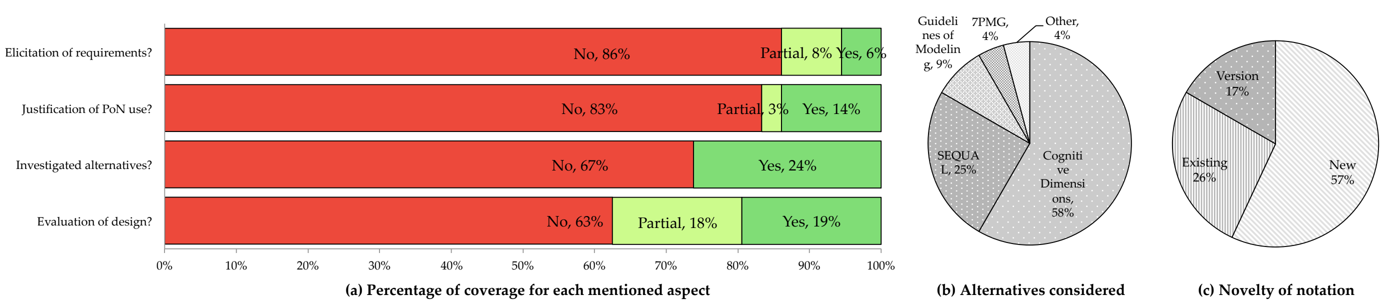


Fig. 5. Extracted data of selected papers. (a) shows for each aspect (elicitation of requirements, justification of PoN use, investigation of alternative approaches, and evaluation of design) the percentage of papers that did not address a factor (-), did so with participants who were not intended users, e.g., students (±), or did so with intended users (+). (b) shows details of investigated alternatives for the 24% of papers that did so. (c) shows whether the application was intended to evaluate an existing notation or to design a new/version of notation..

limitations were discussed, with the authors concluding, "these limitations as well as other disadvantages discussed by Moody support our choice not to base our analysis on CDs." Second, the SEQUAL framework and its perceived limitations were discussed, concluding that SEQUAL's "two main limitations are the level of generality and the lack of theoretical and empirical foundations related to visual aspects of notations." Only after discussing other options for investigating cognitive effectiveness did the authors then present their argument for the PoN that it overcomes the limitations, e.g., "the PoN provides framework that is specifically developed for visual notations," and further justify their use of the PoN by "[the] theory is falsifiable, i.e., the principles can be used to generate predictions, which are empirically testable."

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

RQ3a: To what extent do analyses involve users in determining their requirements for the notation?
Table 3 gives an overview of how many of the selected studies involved users in requirements elicitation. Little to no user involvement for eliciting requirements for the visual notation was found. The exact ratio is shown in Fig. 5(a), coded as for no, ± for claimed, and + for explicitly shown user involvement.

RQ3b: For each principle, to what extent are users involved in operationalizing the principle?
We found no examples of users explicitly being involved in the operationalization of the PoN principles or indeed any type of tailoring of the application. This should not be a surprise given the typical lack of user involvement shown above. Given the criticism directed at the PoN for being vague and ambiguous (see Section 3), this lack of user involvement is particularity noteworthy. Involving users could alleviate some of the ambiguity by grounding the operationalization in user requirements.

RQ3c: To what extent are trade-offs between principles discussed and determined with the involvement of users?
None of the selected papers discussed trade-offs between principles with the involvement of users. This might be explained by the position taken by Moody after publication of the use of the PoN in another widely cited application [40], where the authors proposed that it is more important to achieve overall satisfaction of principles rather than optimize one or more principle(s) to the detriment of others. However, we did find one paper where the authors noted that, in their experience, trade-offs between principles depend on the specific design domain and design issues [41], which stresses the need for user involvement in addressing trade-offs between PoN principles.

RQ3d: Were the design decisions regarding visual elements evaluated such that the measurable improvements in the cognitive effectiveness of the notation, compared to its alternative design, were demonstrated?
Only a small number of authors included an evaluation of their new notation or proposed design changes to an existing notation (see Fig. 5(a)). Papers reporting a re-design of existing notations did also typically not include experiments or observations attempting to establish a baseline, e.g., reading speed, accuracy, and recall of a new notation, but instead focused on user perception of the re-designed notation. Some authors, such as those of Paper 70, evaluated only some aspects of the discussed modeling language, focusing on semantic correctness but not evaluating the visual notation as a whole. For the purposes of our analysis, we did not consider such cases as constituting an evaluation.

RQ3e: Did this evaluation of the designed notation include the intended users of the notation? If not, how do the authors justify their evaluation procedure?
Only a very small number of studies evaluated the impact of their application of the PoN on the cognitive effectiveness of the notation. it is therefore difficult to give a meaningful answer to the question of whether such evaluations included intended users. Nonetheless, as can be seen in Fig. 5(a), where ± indicates an evaluation that was conducted with student participants, the performed evaluations frequently did not include actual or intended users of the notation. Some authors who included in their paper an evaluation with students (e.g., Papers 26, 36, 70) argued explicitly for the validity of evaluation with student participants, citing empirical work providing evidence for students being representative, such as [42], [43].

5.4 How verifiable are the performed analyses?

RQ4a: What is the scope of the analysis in terms of the PoN theory's nine principles?
Table 4 shows which principles were applied in each selected paper, as well as how verifiable this application is. In Fig. 6, the relative number of well reported, claimed (without rationale), ignored, and unverifiable principle applications are shown. It can be seen that some principles, such as cognitive fit, cognitive integration, and complexity

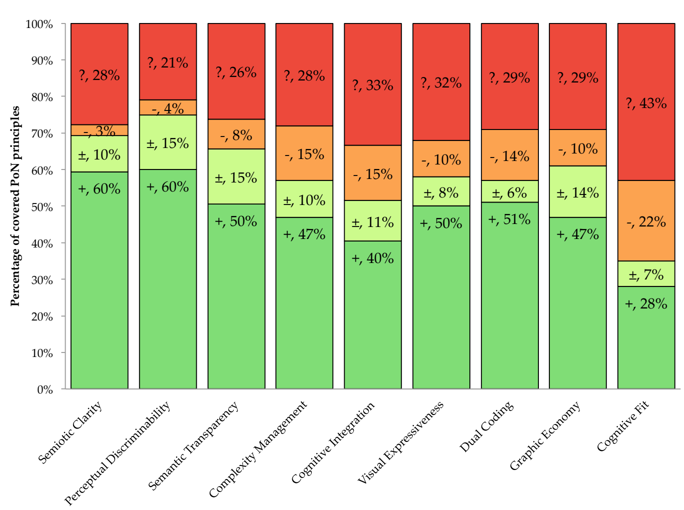


Fig. 6. Scope of addressed principles. + indicates verifiable application, ± indicates claimed application, - indicates claimed irrelevance, and ? indicates unverifiable application.

management, seem to be reported in less detail than other principles.

An interesting difference in the data following from the sensitivity analysis can be seen when these numbers are split into the scope for those studies that used the PoN to analyze an existing notation to assess its cognitive effectiveness versus those that used the PoN in order to, or while, creating a new visual notation or a new version of an existing one. For a clearer overview, we visualize these distributions in Fig. 7, which clearly shows a difference in the number of ? (unverifiable) and, to a lesser degree, ± (claimed but not detailed) classifications between these two types of applications.

The analysis scope of the selected papers can thus be seen to differ according to the exact viewpoint taken (overall or split into types of application). Overall, the authors of most applications seem to treat and report on about half of the PoN principles with some amount of detail. However, if we split the selected papers into those with existing notations and those with new/versions of notations, a clear difference in scope can be seen, with applications of the PoN to new/versions of notations having a markedly smaller scope in terms of detailed and verifiable reporting.

RQ4b: Is the analysis of each principle verifiable?
Given the high number of and ± findings, as evident in Table 4 and Fig. 7, the simple answer is that it is not the case that each principle was analyzed in a systematic, replicable way. Furthermore, even for those principles that were classified as +, there is a large variety in the quality of the application of the principle. However, the reasons for this cannot be attributed solely to the respective authors, but are likely to be related to the PoN itself, as is discussed in more detail in Section 6.3.

Following the sensitivity analysis noted in Sec. 4.3.5, we found a stark difference in terms of the distribution of principle classification between applications of the PoN to design new/versions of notations and applications to evaluate existing notations. Fig. 8 shows these distributions and how they indicate that applications of the PoN to design new/versions of notations are far from an ideal scenario in terms of the verifiability of each principle.

Ideally, Fig. 8(a) would show a left-skewed distribution, that is, all applications would be contained in the right hand side bins, indicating that all applications were explicitly addressed and verifiable. Furthermore, ideally Fig. 8(b)-(d) would present right-skewed distributions, that is, all applications would be in the left hand side bins, indicating that no principle applications were left simply claimed, irrelevant, or unverifiable. The difference between the top and bottom rows show that the application of the PoN to guide the design of new/version of notations is even farther from the ideal scenario than applications that evaluate existing notations.

6 Discussion

6.1 Strengths and Weaknesses

6.1.1 Comparison to other reviews

While there are no other SLRs, or any other reviews with which to compare our results, the scope of PoN applications is wider than mentioned by those few papers that discuss other PoN applications as related work. More importantly, there is an important difference: papers discussing PoN applications typically mention applications of the PoN that evaluated existing notations such as BPMN and WebML. Moody's own analysis of UML that predates the definitive version of the PoN is typically also mentioned, as is sometimes a technical report by Moody analyzing ArchiMate. Only in rare cases are notations actually designed according to the PoN mentioned as related work. Paper 36, for example, mentions the newly designed VTML as an example of a PoN application. This may indicate that authors are not aware of other PoN applications, which limits their ability to compare and contrast how they apply each principle. As a consequence, even though the body of work applying the

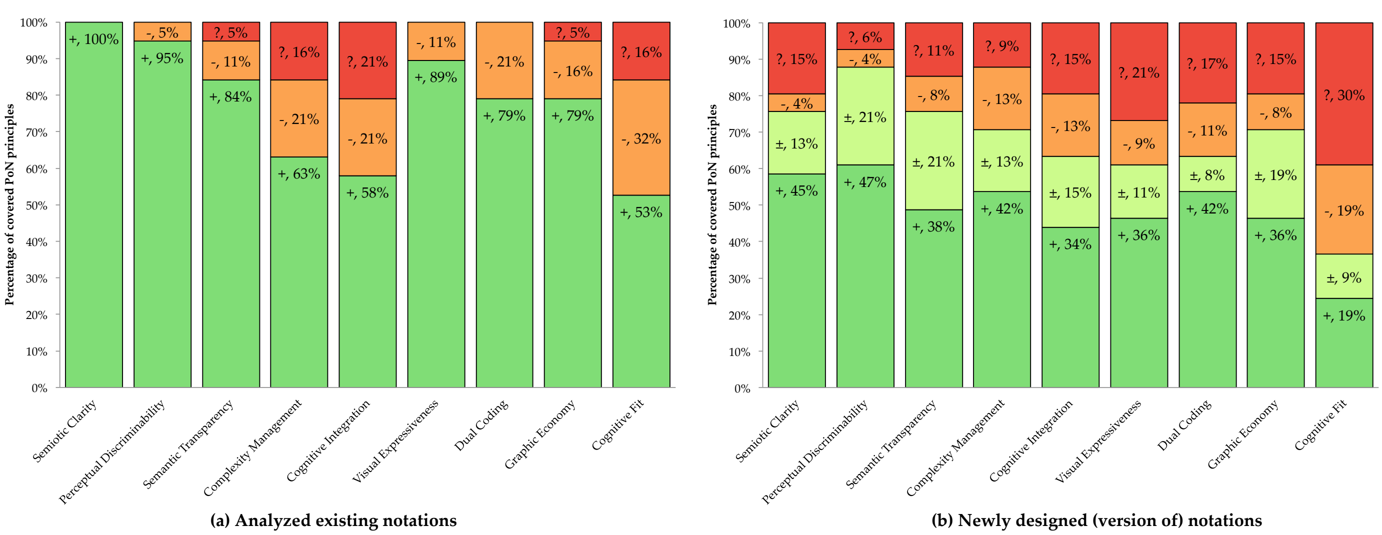


Fig. 7. Scope of addressed principles. As in Fig. 6, + indicates verifiable application, ± indicates claimed application, - indicates claimed irrelevance, and ? indicates unverifiable application. Here, the data are divided into two cohorts; (a) scope of application for applications of the PoN that evaluated existing notations, and (b) scope of application for applications of the PoN that designed new / versions of notations.

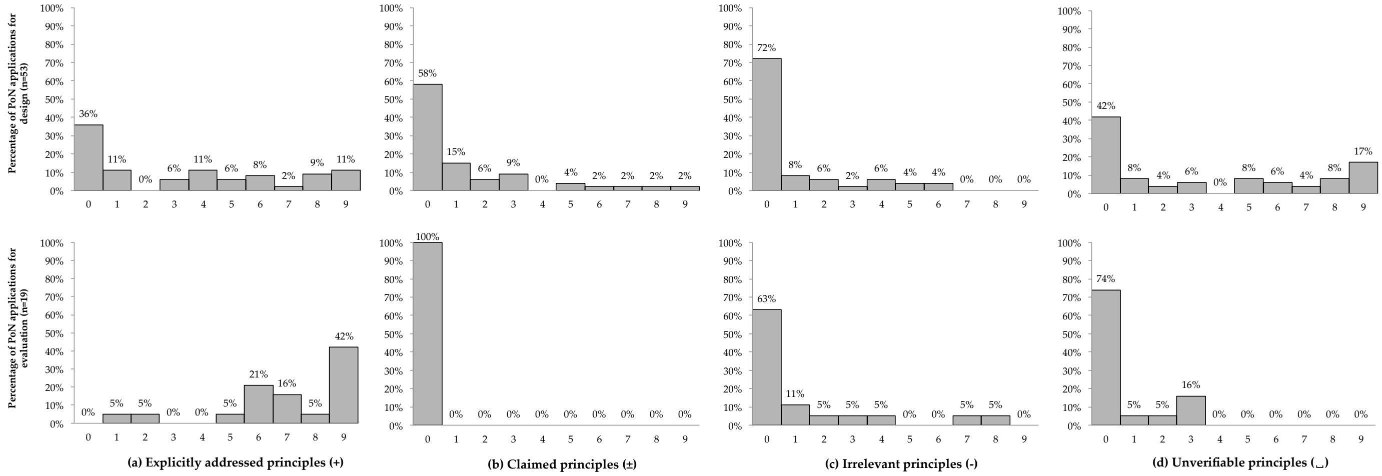


Fig. 8. Comparison of the distribution of principle classifications between applications of the PoN used to evaluate existing notations (bottom row) and to guide the design of new/version of notations (top row).

PoN has grown, this body's use as a source of practical *how* to knowledge has remained limited in practice.

6.1.2 Threats to validity

According to Kitchenham, the most important threats to validity are those that introduce bias into the review process [29]. Zhou et al. further detail the most discussed and prevalent threats to validity of SLRs [44], which we address for this SLR here:

- *Bias in study selection*: as detailed in the review protocol (Sec. 4.3.2), to ensure that we used the inclusion and exclusion criteria effectively and were not guided by subjective conjecture, we documented each decision with quotes from the actual studies. This allows for additional colleagues to verify random samples of inclusion/exclusion decisions in order to assess that we were not guided by subjective opinion.
- *Bias in data extraction*: we discussed the use of coding schemes for data, which had to be coded (e.g., requirements, justification, evaluation, alternatives), as well as the principle scope analysis. The utilized coding scheme was documented and used during any discussions of different decisions between the authors to ensure no bias or misinterpretation of the coding occurred.
- *Inappropriate or incomplete search terms in automatic search*: the procedure for determining the search string (see Sec. 4.3.1) was based on iterative refinement using identified paper examples as input. Additionally, we used multiple databases to ensure full coverage, as well as forward snowballing on the PoN theory's definitive article and validation of the completeness of the selection with Google Scholar.

Furthermore, this review might be threatened by some additional, more specific, validity considerations:

- *Lack of standard languages and terminologies*: studies do not consistently use the same term to refer to visual notations. Terms like notation, diagram, and modeling are all used. This was primarily a difficulty while building the search string. For the inclusion stage, we followed Kitchenham's guidelines and applied the second inclusion criterion liberally, so as to avoid missing any studies because of unexpected use of terminology. Furthermore, the reading of foreign language articles could have posed a threat, but all retrieved articles had an English abstract and keywords, and thus, this threat was avoided.
- *Paper/database inaccessible*: we ran into several cases where we could not access papers because of more niche publishers to which we or our immediate colleagues had no access. All these cases were resolved by (1) requesting a full-text through ResearchGate and/or Academia.edu, (2) if no response, emailing the first author, and (3) if no response, emailing remaining authors. Following this approach, we were able to acquire full texts of every study retrieved by the search query.
- *Primary study duplication*: we took care to avoid including redundant applications of the PoN per our exclusion criterion. When a duplication was found, we considered which paper offered the most details about the PoN application, and included only that one in the review. We did so because in some cases later papers focused more on other aspects, e.g., implementation of a tool or case studies, and omitted details of the visual notation's design included in older papers.

6.2 Meaning of findings: on the PoN's actual impact

Increasing use for designing new notations

Fig. 9 shows that the PoN is increasingly used for designing new notations, whereas Figs. 7 and 8 show there is a lot of effort to be invested in applications to new notations. The fact that the PoN is increasingly used to design new notations instead of evaluating existing notations places more stress on ensuring it is well applied. For a design theory like the PoN that means ensuring verifiability and design rationale is important.

Lack of justification

As much as the use of the PoN has grown over the years, there is very little justification by authors as to why the PoN theory was chosen as the approach to ensure the visual

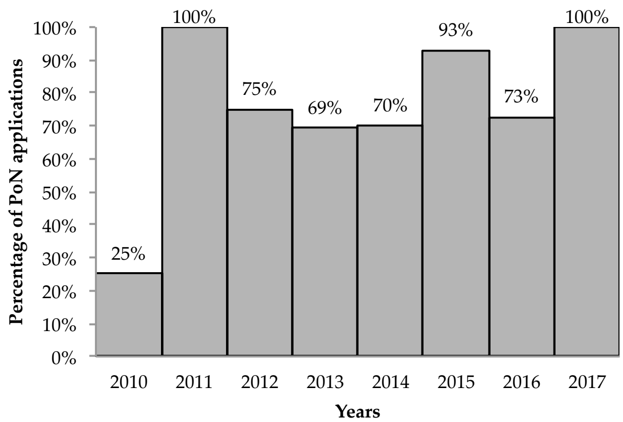


Fig. 9. Histogram showing frequency of PoN applications used to design new notations since 2010.

notation's cognitive effectiveness. This indicates, perhaps, that the selection of the PoN theory is driven by its wider acceptance of those in the field, signaling that it is an expected part of creating a new notation. However, this runs the risk of leading to scenarios where little more than lip service is paid to the PoN theory, which can already be found in examples below.

Whereas other approaches such as SEQUAL, GoM, or CD are described as "frameworks" or "guidelines," the PoN sets itself apart by clearly identifying itself as a *theory* for visual notation design. Several authors ground use of the PoN specifically in terms of its being scientific and providing a scientific means. This seems to be inspired by the PoN's main claim of the need to ground design rationale in *scientific evidence* rather than subjective opinion. However, given the lack of experimental evaluation or otherwise rigorous testing of proposed design, what authors understand as "scientific" remains vague. For example, in Paper 36 the PoN is characterized as being a much-awaited "welcome theory," because unlike its predecessors it makes it possible "to conduct notation evaluations scientifically." In Paper 16, it is similarly stated that "the PoN principles are applied because of their scientific and theoretical validity." In Paper 26, the title clearly claims "A scientific evaluation of [. . .]" Other approaches, in particular CD, are ruled out in this paper because of "the limitations of CDs framework with respect to being a scientific basis for evaluating and designing visual notations [. . .]" Given the authors' further recommendation for more analyses of visual notations, which ". . . should be supplemented with improvements that use the PoN principles as a scientific basis" (emphasis added), the PoN's suitability seems to be asserted because of its nature as a scientific theory.

This re-affirmation of the PoN as a scientific theory and its value as compared to competing or complementary approaches imposes requirements on how the PoN is used. A high level of rigor in all stages of research should be expected, but may not be that easily achieved, given the difficulties in applying the PoN, as discussed in Section 2.

Lack of user involvement

The finding that there is typically a lack of user involvement is especially troubling, given the need to involve users in operationalizing principles. This indicates that the PoN is not as well applied as it should be: how can a notation be designed to be semantically transparent for its intended audience, if members of this audience were not involved in either the notation design, or its evaluation?

In a recent study, requirements for a process modeling notation were elicited, and the authors found they had a strong overlap with the PoN principles: "Interestingly a lot 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.' Furthermore, demanding non-redundant symbolic corresponds to the Principle of Semiotic Clarity."' [45]

If it is indeed found that practitioners have requirements of importance to them that are strongly related to the PoN principles, then they should definitely be involved in the operationalization of the principles for any application, as well as a later evaluation of the actually designed visual notation.

Lack of comparability because of application scope

Figs. 6 and 7 show that rarely are all principles explicitly addressed. The PoN's theoretical model includes nine independent variables (the principles) that affect the dependent variable (cognitive effectiveness), with the claim that visual notations that satisfy the principles are more cognitively effective than those that do not. However, if only a partial number of principles are applied, or some principles are only partially applied, how can the degree to which notations are more cognitively effective be clearly established?

For example, some authors [46] selected a limited number of PoN principles according to earlier empirical work that examined which principles were perceived to be most useful by their users [47].

As a result, it is not necessarily feasible to compare two visual notations for similar purposes, which were designed according to the PoN, and pick the most cognitively effective one. On the other hand, given the trade-offs between principles, if the exact principles applied were well described and verifiable, one might be able to select a visual notation that best satisfies the exact requirements for a specific audience and purpose.

Lack of verifiability of principle application

The most critical finding of this review is that for many newly designed visual notations it cannot be verified whether certain principles hold (see Fig. 7(b)). This is a serious issue, because it impacts the certainty with which we can say that the PoN theory was actually used, let alone satisfies its main message that design should be grounded in (scientific) evidence rather than subjective opinion. Fig. 8 further shows that this lack of verifiability is most evident in the applications of the PoN used to design new/versions of notations, even though here it is most important to ensure the well-foundedness of the notation's design rationale.

Let us look in some more detail how this actually manifests in different PoN applications. The authors of some papers in which a new notation was introduced claim to have used the PoN or its principles, but provide no further details of how these principles guided the design or how the created design satisfies the principles. Some other authors explicitly state the principles of the PoN that they used or followed, but provide no further information. For example, in Paper 70 it is stated that "... the design of the eCRG language partially considers the principles for designing effective visual notations. Particularly, the concepts of semiotic clarity, perceptual discriminability, semantic transparency, graphic economy, and cognitive fit were taken into account." However, in the remainder of the paper no further reference to or assessment of these principles was made, making it impossible to verify the degree to which the visual notation was actually designed with these principles in mind, let alone satisfied them.

Another example is found in Paper 5, where the authors first claim that all the principles of the PoN were used as guidelines in the design of the visual notation: "The notation follows the principles of Moody's prescriptive theory for visual notations..." However, the following discussion does not link this statement to the exact terminology of the PoN, making it difficult to assess what was actually done or which principles were (meant to be) addressed. For example, when discussing the design of symbolic representations, the authors state "the hardest part in constructing adequate icons is to find the simplest symbol, which sends the clearest message in a unique and easy way." This can be taken to refer to semantic transparency (sending clear messages), but also to perceptual discriminability and visual expressiveness (a unique way, i.e., a unique combination of visual variables.)

Some authors clearly note which principles of the PoN they selected and which were deferred to future work, but still make claims as to the remaining principles' satisfaction. For example, in Paper 33, although the authors note that they did not want to make any claims about whether some principles were or were not satisfied, because further validation was required, they still claim (partial) satisfaction with no data, e.g., of semantic transparency: "[...] the visual metaphors applied to each of the elements are representative, and its semantic is the closest possible to its intended meaning. However, validating this principle also requires an experimental study."

Additionally, some authors do not refer to the theory or data in which some of their assumptions are grounded. For example, in Paper 2 in the context of satisfying graphic economy, the authors state "[w]e also have that the number of different graphical symbols in the model is under the upper limit of six categories for graphics complexity, so the principle of graphic economy is accomplished." The graphic economy principle indeed prescribes that a number of distinct graphical symbols greater than some threshold n should not be used, for which multiple examples and sources are given. The particular threshold given here, while probably meant to refer to Miller's Law of 7 ± 2 [48], is given arbitrarily as six, with no reference to either an original theory or other work reporting its use. Thus, it cannot be verified whether the design rationale indeed derives from scientific evidence, or is a subjective opinion, here a "lucky shot."

A more positive example of a verifiable application of the PoN can be found in, for example, Paper 4. The authors clearly state whether PoN principles were satisfied or not, and how. For example, for semiotic clarity: "To that end, our visual notation uses a different symbol to represent each of the taxonomy's elements, and does not introduce any new concepts." This statement is straightforwardly verified by the explicit inclusion of extensive meta-models and tables detailing each visual element of the visual notation. Thus, readers can immediately see and verify that this principle indeed holds.

6.3 Is it the PoN, its followers, or both?

The previous subsection discussed *what* is troubling about the studied PoN applications. But a more pressing question might be, *why*, or even more direct, because of *whom* are these applications so troubling? Are these troubles to be attributed solely to those applying the PoN, the PoN itself, or both?

On the one hand, not all these challenges can be attributed solely to the people applying the PoN. Working with the PoN principles is further confounded by the difficulty of determining the extent to which a principle is satisfied, as noted by some authors, e.g., in Paper 16: "There is no distinct method for measuring the extent to which the criteria are fulfilled." [49] The criticism directed toward the PoN in the literature, as discussed previously, shows indeed that, as a theory, it does not always offer exact guidance to ensure that PoN users can apply it systematically and verifiably.

On the other hand, in some cases authors also seem to invest less effort than ideally required to satisfy PoN principles. For example, in Paper 24 the authors present arguments that allow a debate as to whether they satisfy a principle or merely pay lip service to it. For example, for dual coding, the authors stated "All SEAM [the proposed notation] graphical symbols are accompanied by short labels or descriptive textual expressions." [39] Yet, this does not say anything about the quality or cognitive impact of the exact text, for example, even though there are discussions in the literature providing some more substance in this context. For example, while the PoN does not explicitly enforce it, Moody et al. set an arguably higher standard for dual coding in the PoN's application to i* [40]. Their recommendation includes to not only ensure all elements are named, but that there are clear guidelines for doing so. For example, they recommend using a *verb-object form* to name i* elements, and recommend avoiding confusing users by encoding roles with too similar role names such as *depender* versus *dependee*. In fact, they raise the issue that "the meaning of those terms would not be understood by the average business user as they are not commonly used in everyday language." [40] It thus seems apparent that dual coding should involve more consideration than a mere binary check of whether text is used in addition to graphics.

Several authors used arguments for the satisfaction of a particular principle by referring to examples of other notations that would not satisfy that principle. For example, in Paper 33, the authors presented their claims for a number of principles and determined that their notation satisfies graphic economy, arguing that the proposed notation "[...] uses a maximum of 12 graphical symbols for each diagram. This is a small number compared to UML Class Diagrams that have a graphic complexity of over 40." [50] While Moody and Van Hillegersberg claimed that the graphic complexity of UML Class Diagrams "exceeds human discrimination ability by quite a large margin," [6] using such forms of ad hominem does not answer the question whether the notation investigated actually satisfies the principle. Noting counter-examples does not substitute for design rationale, and should not be used as a sole argument for whether a particular principle holds.

Finally, in Paper 26 another case exists where the authors, while not relying solely on it, present an argument to authority for the validity of their analysis by explicitly naming and showing that the original notation's designer agreed with their assessment and redesign: "One notable survey respondent is Dr. Andreas Opdahl, one of the original creators of the notation. Dr. Andreas Opdahl correctly answered all questions in the first section of the questionnaire and has preferred the entire set of the new notation symbols." [51] Such arguments should be avoided at all costs, because they draw attention away from the design rationale itself, potentially biasing the reader to bypass direct examination of the presented evidence, trusting instead in the supposed expertise of an authority [52].

A different type of potential fallacy can be found, for example, in Paper 1, where semantic transparency is argued for because "An AGENT could be shown wearing dark glasses and holding a gun (by association with agents of the 007 kind)" [40]. In such cases, the authors run into a potential anecdotal fallacy, arguing from their own experience and cultural frame of reference, without explicating that those statements would hold only for people with the same reference frame. The importance of considering such factors can be seen in recent studies on IS, which have shown culture-specific differences of understanding for color-coding in process models, and proposed cultural adaptive model design [53].

While it has been argued that the PoN has its share of challenges, is difficult to apply, and does not offer enough explicit guidance, we cannot avoid re-iterating that many authors seem to have forgotten or misunderstood the PoN's main message in the article publishing it, indicated in the abstract, at the start of the article, and re-iterated in the conclusions: that design rationale "should be based on scientific evidence rather than subjective criteria, as is currently the case." [8]

7 Conclusion

7.1 Contributions

This article presented an SLR of the applications of a theory for validating the cognitive effectiveness of visual notations, the Physics of Notations. We investigated the purposes for which the PoN has been applied in the recent years following its publication, investigating which visual notations have been designed or analyzed using it, looking in particular at the verifiability and completeness of these applications.

While the PoN has gained traction among the research community, as evidenced by the 70 applications, only a limited number of those applications can be said to be in line with the PoN's main message on the importance of explicit design rationale, grounded in scientific evidence. The contribution of this review, thus, is that it serves as an indicator of the current gap between the vision promoted by the PoN for visual notation design, and the reality. There are multiple reasons for this gap, stemming both from the PoN itself and from its followers.

With regard to the PoN's followers, this review has shown that many authors of applications simply claim that principles have been applied, while offering little to no evidence or supporting design rationale. Furthermore, there is a lack of user involvement, and many applications differ in scope. As a result, the landscape of PoN applications is fragmented, in effect devaluating the impact of the PoN itself. With little possible comparison between individual applications, and the general uncertainty of what exactly was done, PoN applications that can serve other researchers to learn by example are few and far between.

With regard to the PoN itself, another, perhaps more fundamental, concern is to what extent the PoN supports researchers in systematically applying each of the PoN principles and verifying whether they hold. Some of the criticism leveled at applications of the PoN may indeed be rooted in the different degrees to which each principle lends itself to be operationalized. Researchers are required to interpret those principles, which may be difficult to operationalize, and then determine exactly what to do. This may lead to many different possible outcomes. While the PoN, as it is, represents one of the premier approaches for assessing the cognitive effectiveness of visual notations, the manner in which it is applied does not reach its full potential.

7.2 Recommendations

Interpreting our findings, there are two main aspects to improving the value of the PoN theory and ensuring that it actually leads to improved design of visual notations: (1) those who apply the PoN need to take more care to ensure their application is in line with the PoN's main message, in particular to stress the evidence for any design rationale given, and (2) support for use of the PoN needs to offer more explicit guidance for applying the actual principles, in particular those requiring user involvement for operationalization and evaluation.

With regard to (1), *applying the PoN* should be done with more care. Those using the PoN to design notations should pay additional attention to principles requiring considerate application of cognitive theories, such as *dual coding, cognitive fit, cognitive integration*. In implementing these principles, more than the bare consideration should be given, some points of which have been discussed in Sec. 2. Furthermore, applications of principles (additionally) requiring user involvement in their operationalization and evaluation should always explicitly do so. Design rationale based on expert opinion is no substitute for design rationale inspired by actual users' requirements, and shaped by their feedback in evaluation, iteratively if necessary.

In regard to (2), *improving the PoN* should focus on providing means to guide researchers in applying individual principles, and in doing so enforce the systematic documentation of design rationale and evidence. While tools such as CEViNEdit [26] and approaches such as PoN-S [23] are a welcome start to help structure the process of developing visual notations, more practical guidance on the actual contents of the principles is needed. The work of Störrle and Fish [24] is valuable here, but remains limited to the subset of principles that lend themselves to be formally captured. For those other principles requiring consideration of cognitive theories and user involvement, approaches other than formalization may be needed. For example, structured guidelines can be provided to assist researchers and designers of notations in the elicitation of data for specific principles and categorization of the elicited evidence.

These recommendations stand, hopefully, to lead to a landscape of PoN applications that is of higher value to both researchers and intended users of notations. With clearer scope and explicit design rationale, intended users could more accurately select a suitable notation that fits their intrinsic requirements [45]. Researchers, at the same time, will have a body of notations better grounded in design rationale and evidence to build on for further work. More importantly, this body of notations can effectively serve as instruction material for future applications of the PoN, alleviating at least somewhat the established difficulty of applying the PoN.

REFERENCES

[1] Y. Wand and R. Weber, "Research commentary: information systems and conceptual modelinga research agenda," *Information Systems Research*, vol. 13, no. 4, pp. 363–376, 2002.

[2] A. van Lamsweerde, *Requirements Engineering: From System Goals to UML Models to Software Specifications*. Wiley Publishing, 2009.

[3] J. Mylopoulos, "Conceptual modelling and telos 1," *Conceptual modeling, databases, and CASE*, pp. 363–376, 2008.

[4] T. DeMarco, *Structure analysis and system specification*. Springer, 1979, pp. 255–288.

[5] J. H. Larkin and H. A. Simon, "Why a diagram is (sometimes) worth ten thousand words," *Cognitive science*, vol. 11, no. 1, pp. 65–100, 1987.

[6] D. L. Moody and J. van Hillegersberg, "Evaluating the visual syntax of uml: An analysis of the cognitive effectiveness of the uml family of diagrams," in *Software Language Engineering*. Springer, 2008, pp. 16–34.

[7] N. H. Narayanan and R. Hübscher, *Visual language theory: Towards a human-computer interaction perspective*. Springer, 1998, pp. 87–128.

[8] 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.

[9] S. Hitchman, "The details of conceptual modelling notations are important-a comparison of relationship normative language," *Communications of the Association for Information Systems*, vol. 9, no. 1, p. 10, 2002.

[10] D. van der Linden and I. Hadar, "Evaluating the evaluators," in *Proceedings of the 11th International Conference on Evaluation of Novel Software Approaches to Software Engineering*, ser. ENASE 2016. SCITEPRESS - Science and Technology Publications, Lda, 2016, pp. 222–227.

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

[12] N. Burlutskiy, M. Petridis, A. Fish, and N. Ali, "How to visualise a conversation: case-based reasoning approach," 2014.

[13] K. Doore and P. Fishwick, "Prototyping an analog computing representation of predator prey dynamics," in *Simulation Conference (WSC), 2014 Winter*, 2014, pp. 3561–3571.

[14] C. H. Kim, J. Grundy, and J. Hosking, "A suite of visual languages for model-driven development of statistical surveys and services," *Journal of Visual Languages & Computing*, vol. 26, pp. 99–125, 2015.

[15] L. Vickery, "Notational semantics in music visualization and notation," in *Australasian Computer Music Conference 2014*, 2014, p. 101.

[16] D. Naranjo, M. Sanchez, and J. Villalobos, "Visual analysis of enterprise models," in *Enterprise Distributed Object Computing Conference Workshops (EDOCW), IEEE 16th International*, 2012, pp. 19–28.

[17] M. Cortes-Cornax, S. Dupuy-Chessa, D. Rieu, and N. Mandran, "Evaluating the appropriateness of the bpmn 2.0 standard for modeling service choreographies: using an extended quality framework," *Software & Systems Modeling*, vol. 15, no. 1, pp. 219–255, 2016.

[18] C. Natschläger, F. Kossak, and K.-D. Schewe, "Deontic bpmn: a powerful extension of bpmn with a trusted model transformation," *Software & Systems Modeling*, vol. 14, no. 2, pp. 765–793, 2013.

[19] E. R. Luna, G. Rossi, and I. Garrigos, "Webspec: a visual language for specifying interaction and navigation requirements in web applications," *Requirements Engineering*, vol. 16, no. 4, pp. 297–321, 2011.

[20] B. Henderson-Sellers, "Towards the consolidation of a diagramming suite for agent-oriented modelling languages," *ISRN Software Engineering*, vol. 2013, 2013.

[21] D. van der Linden, A. Zamansky, and I. Hadar, "How cognitively effective is a visual notation? on the inherent difficulty of operationalizing the physics of notations," in *Enterprise, Business-Process and Information Systems Modeling: 17th International Conference, BPMDS 2016, 21st International Conference, EMMSAD 2016, Held at CAiSE 2016, Ljubljana, Slovenia, June 13-14,2016 , Proceedings*. Springer, 2016, pp. 448–462.

[22] J. Gulden and H. A. Reijers, "Toward advanced visualization techniques for conceptual modeling," in *Proceedings of the CAiSE Forum*, 2015, pp. 8–12.

[23] M. d. G. da Silva Teixeira, G. K. Quirino, F. Gailly, R. de Almeida Falbo, G. Guizzardi, and M. P. Barcellos, "PoN-S: A systematic approach for applying the physics of notation (pon)," in *Enterprise, Business-Process and Information Systems Modeling: 17th International Conference, BPMDS 2016, 21st International Conference, EMMSAD 2016, Held at CAiSE 2016, Ljubljana, Slovenia, June 13-14,2016 , Proceedings*. Springer, 2016, pp. 432–447.

[24] H. Störrle and A. Fish, "Towards an operationalization of the "Physics of Notations" for the analysis of visual languages," in *International Conference on Model Driven Engineering Languages and Systems*. Springer, 2013, pp. 104–120.

[25] J. Stark, R. Braun, and W. Esswein, "Systemizing colour for conceptual modeling," in *Wirtschaftsinformatik 2017*, 2017.

[26] D. Granada, J. M. Vara, V. A. Bollati, and E. Marcos, "Enabling the development of cognitive effective visual DSLs, booktitle = International Conference on Model Driven Engineering Languages and Systems, publisher = Springer, year = 2014, pages = 535-551."

[27] D. van der Linden and I. Hadar, "User involvement in applications of the pon," in *Lecture Notes in Business Information Processing*, vol. 249. Springer, pp. 109–115.

[28] D. van der Linden, A. Zamansky, and I. Hadar, "A framework for improving the verifiability of visual notation design grounded in the physics of notations," in *Requirements Engineering Conference (RE), 2017 IEEE 25th International*. IEEE, 2017, pp. 41–50.

[29] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," EBSE Technical Report EBSE-2007-01, 2007.

[30] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering–a systematic literature review," *Information and software technology*, vol. 51, no. 1, pp. 7–15, 2009.

[31] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineeringa tertiary study," *Information and Software Technology*, vol. 52, no. 8, pp. 792–805, 2010.

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

[33] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, and G. Pappas, "Comparison of pubmed, scopus, web of science, and google scholar: strengths and weaknesses," *The FASEB journal*, vol. 22, no. 2, pp. 338–342, 2008.

[34] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of systems and software*, vol. 80, no. 4, pp. 571–583, 2007.

[35] H. G. Frankfurt and M. Bischoff, *On bullshit*. Cambridge Univ. Press, 2005.

[36] J. Krogstie, G. Sindre, and H. Jørgensen, "Process models representing knowledge for action: a revised quality framework," *European Journal of Information Systems*, vol. 15, no. 1, pp. 91–102, 2006.

[37] R. Schuette and T. Rotthowe, "The guidelines of modelingan approach to enhance the quality in information models," in *International Conference on Conceptual Modeling*. Springer, 1998, pp. 240–254.

[38] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst, "Seven process modeling guidelines (7pmg)," *Information and Software Technology*, vol. 52, no. 2, pp. 127–136, 2010.

[39] G. Popescu and A. Wegmann, "Using the physics of notations theory to evaluate the visual notation of seam," in *IEEE 16th Conference on Business Informatics (CBI)*, vol. 2, 2014, pp. 166–173.

[40] D. L. Moody, P. Heymans, and R. Matulevičius, "Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation," *Requirements Engineering*, vol. 15, no. 2, pp. 141–175, 2010.

[41] J. C. Thomas, J. Diament, J. Martino, and R. K. E. Bellamy, "Using the "physics" of notations to analyze a visual representation of business decision modeling," in *Visual Languages and Human-Centric Computing (VL/HCC), 2012 IEEE Symposium on*, 2012, pp. 41–44.

[42] M. Höst, B. Regnell, and C. Wohlin, "Using students as subjects–a comparative study of students and professionals in lead-time impact assessment," *Empirical Software Engineering*, vol. 5, no. 3, pp. 201–214, 2000.

[43] M. Svahnberg, A. Aurum, and C. Wohlin, "Using students as subjects-an empirical evaluation," in *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*. ACM, 2008, pp. 288–290.

[44] X. Zhou, Y. Jin, H. Zhang, S. Li, and X. Huang, "A map of threats to validity of systematic literature reviews in software engineering," in *Software Engineering Conference (APSEC), 2016 23rd Asia-Pacific*. IEEE, 2016, pp. 153–160.

[45] J. Wiebring and K. Sandkuhl, *Selecting the "Right" Notation for Business Process Modeling: Experiences from an Industrial Case*. Springer, 2015, pp. 129–144.

[46] J. M. Netto, J. B. S. Franca, F. A. Baiao, and F. M. Santoro, "A notation for knowledge-intensive processes," in *IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2013, pp. 190–195.

[47] K. Figl and M. Derntl, "The impact of perceived cognitive effectiveness on perceived usefulness of visual conceptual modeling languages," in *International Conference on Conceptual Modeling*. Springer, 2011, pp. 78–91.

[48] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological review*, vol. 63, no. 2, p. 81, 1956.

[49] J. Herter, R. Brown, and J. Ovtcharova, "A visual language for the collaborative visualization of integrated conceptual models in product development scenarios," in *Smart Product Engineering*. Springer, 2013, pp. 805–814.

[50] J. J. B. García, A. C. C. Ramos, and J. A. P. Mariscal, "Towards a domain-specific language to design adaptive software: the dmlas approach," *Ingenieria y Universidad*, vol. 20, no. 2, 2016.

[51] F. Saleh and M. El-Attar, "A scientific evaluation of the misuse case diagrams visual syntax," *Information and Software Technology*, vol. 66, pp. 73–96, 2015.

[52] L. Cummings, *Argument from Authority*. Springer, 2015, pp. 67–92.

[53] T. F. Kummer, J. Recker, and J. Mendling, "Enhancing understandability of process models through cultural-dependent color adjustments," *Decision Support Systems*, vol. 87, pp. 1–12, 2016.

[54] E. Martinez, G. Diaz, M. E. Cambronero, and G. Schneider, "A model for visual specification of e-contracts," in *IEEE International Conference on Services Computing (SCC)*, 2010, pp. 1–8.

[55] K. Figl, M. Derntl, M. C. Rodriguez, and L. Botturi, "Cognitive effectiveness of visual instructional design languages," *Journal of Visual Languages & Computing*, vol. 21, no. 6, pp. 359–373, 2010.

[56] P. Laurent, P. Mader, J. Cleland-Huang, and A. Steele, "A taxonomy and visual notation for modeling globally distributed requirements engineering projects," in *5th IEEE International Conference on Global Software Engineering (ICGSE)*, 2010, pp. 35–44.

[57] J. Helming, M. Koegel, F. Schneider, M. Haeger, C. Kaminski, B. Bruegge, and B. Berenbach, "Towards a unified requirements modeling language," in *Requirements Engineering Visualization (REV), 2010 Fifth International Workshop on*. IEEE, 2010, pp. 53–57.

[58] N. Genon, P. Heymans, and D. Amyot, *Analysing the cognitive effectiveness of the BPMN 2.0 visual notation*. Springer, 2011, pp. 377–396.

[59] N. Genon, D. Amyot, and P. Heymans, *Analysing the cognitive effectiveness of the UCM visual notation*. Springer, 2011, pp. 221–240.

[60] P. Etcheverry, C. Marquesuza, T. Nodenot *et al.*, "A visual programming language for designing interactions embedded in web-based geographic applications," in *ACM International Conference on Intelligent User Interfaces*, 2012, pp. 207–216.

[61] M. Famelis and S. Santosa, "Mav-vis: a notation for model uncertainty," in *Modeling in Software Engineering (MiSE), 2013 5th International Workshop on*. IEEE, 2013, pp. 7–12.

[62] B. Weisenberger and B. Vogel-Heuser, "Evaluation of a graphical modeling language for the specification of manufacturing execution systems," in *IEEE 17th Conference on Emerging Technologies & Factory Automation (ETFA)*, 2012, pp. 1–7.

[63] R. Matulevicius, H. Mouratidis, N. Mayer, E. Dubois, and P. Heymans, "Syntactic and semantic extensions to secure tropos to support security risk management," *Journal of Universal Computer Science*, vol. 18, no. 6, pp. 816–844, 2012.

[64] U. Breitenbücher, T. Binz, O. Kopp, F. Leymann, and D. Schumm, *Vino4TOSCA: A visual notation for application topologies based on TOSCA*. Springer, 2012, pp. 416–424.

[65] Z. Dobesova, "Using the physics of notation to analyse model-builder diagrams," *Proceedings of SGEM*, pp. 595–602, 2013.

[66] I. Plauska and R. Damaševičius, *Design of Visual Language Syntax for Robot Programming Domain*. Springer, 2013, pp. 297–309.

[67] S. Boone, M. Bernaert, B. Roelens, S. Mertens, and G. Poels, *Evaluating and improving the visualisation of CHOOSE, an enterprise architecture approach for SMEs*. Springer, 2014, pp. 87–102.

[68] C. Miske, M. A. Rothenberger, and K. Peffers, "Towards a more cognitively effective business process notation for requirements engineering," in *Desrist*, 2014, pp. 360–367.

[69] S. Anwer and M. El-Attar, "An evaluation of the statechart diagrams visual syntax," in *International Conference on Information Science and Applications (ICISA)*, 2014, pp. 1–4.

[70] J. Michael and H. C. Mayr, *Conceptual modeling for ambient assistance*. Springer, 2013, pp. 403–413.

[71] M. Malinova, "A language for process map design," in *Business Process Management Workshops*, pp. 567–572.

[72] E. Woods and R. Bashroush, "Modelling large-scale information systems using adls-an industrial experience report," *Journal of Systems and Software*, vol. 99, pp. 97–108, 2015.

[73] M. Hess, M. Kaczmarek, U. Frank, L. Podleska, and G. Tger, "A domain-specific modelling language for clinical pathways in the realm of multi-perspective hospital modelling," in *ECIS*, 2015.

[74] D. Romuald and T. Mens, "Gismo: a domain-specific modelling language for executable prototyping of gestural interaction," in *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, 2015, pp. 34–43.

[75] B. Roelens and G. Poels, "The development and experimental evaluation of a focused business model representation," *Business & Information Systems Engineering*, vol. 57, no. 1, pp. 61–71, 2015.

[76] D. J. Rough and A. J. Quigley, "Jeeves–a visual programming environment for mobile experience sampling," in *Visual Languages and Human-Centric Computing (VL/HCC), 2015 IEEE Symposium on*. IEEE, 2015.

[77] D. L. Rodrigues, F. M. Santoro, F. A. Baiao, and J. M. Netto, "Towards a context-based representation of the dynamicity perspective in knowledge-intensive processes," in *IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2015, pp. 165–170.

[78] J. Liu, J. Grundy, I. Avazpour, and M. Abdelrazek, "A domain-specific visual modeling language for testing environment emulation," in *2016 IEEE Symposium on Visual Languages and Human-Centric Computing*. IEEE, 2016.

[79] U. Breitenbücher, P. Hirmer, K. Kpes, O. Kopp, F. Leymann, and M. Wieland, "A situation-aware workflow modelling extension,"

in *International Conference on Information Integration and Web-based Applications & Services*. ACM, 2015, p. 64.

[80] M. Hanser, C. Di Ciccio, and J. Mendling, "A novel framework for visualizing declarative process models," in *ZEUS*. CEUR-WS, 2016, p. 5.

[81] M. Saeed, F. Saleh, S. Al-Insaif, and M. El-Attar, "Empirical validating the cognitive effectiveness of a new feature diagrams visual syntax," *Information and Software Technology*, vol. 71, pp. 1–26, 2016.

[82] P. Mäder and J. Cleland-Huang, "A visual language for modeling and executing traceability queries," *Software and Systems Modeling*, vol. 12, no. 3, p. 537, 2013.

[83] N. Elouali, X. Le Pallec, J. Rouillard, and J.-C. Tarby, "A model-based approach for engineering multimodal mobile interactions," in *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*. ACM, 2014, pp. 52–61.

[84] A. S. Algablan and S. S. Somé, "A visual syntax for larman's operation contracts," in *International Conference on Engineering & MIS (ICEMIS)*. IEEE, 2016, pp. 1–9.

[85] Z. Dangarska, K. Figl, and J. Mendling, "An explorative analysis of the notational characteristics of the decision model and notation (dmn)," in *Enterprise Distributed Object Computing Workshop (EDOCW), 2016 IEEE 20th International*. IEEE, 2016, pp. 1–9.

[86] J. Van den Bergh, K. Luyten, and K. Coninx, "Cap3: context-sensitive abstract user interface specification," in *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 2011, pp. 31–40.

[87] K. Sandkuhl and H. Ko, "Component-based method development: an experience report," in *IFIP Working Conference on The Practice of Enterprise Modeling*. Springer, 2014, pp. 164–178.

[88] S. Strohmeier and F. Röhrs, "Conceptual modeling in human resource management: A design research approach," *AIS Transactions on Human-Computer Interaction*, vol. 9, no. 1, pp. 34–58, 2017.

[89] R. Zhang, J. Hosking, J. Grundy, N. Mehandjiev, and M. Carpenter, "Design of a suite of visual languages for supply chain specification," in *Visual Languages and Human-Centric Computing (VL/HCC), 2010 IEEE Symposium on*. IEEE, 2010, pp. 240–243.

[90] S. Gopalakrishnan and G. Sindre, "Diagram notations for mobile work processes," in *IFIP Working Conference on The Practice of Enterprise Modeling*. Springer, 2011, pp. 52–66.

[91] Y. El Ahmar, S. Gérard, C. Dumoulin, and X. Le Pallec, "Enhancing the communication value of uml models with graphical layers," in *ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS)*. IEEE, 2015, pp. 64–69.

[92] F. B. Aydemir, P. Giorgini, J. Mylopoulos, and F. Dalpiaz, "Exploring alternative designs for sociotechnical systems," in *IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2014, pp. 1–12.

[93] M. El-Attar, H. Luqman, P. Karpati, G. Sindre, and A. L. Opdahl, "Extending the uml statecharts notation to model security aspects," *IEEE Transactions on Software Engineering*, vol. 41, no. 7, pp. 661–690, 2015.

[94] T. Kühn, K. Bierzynski, S. Richly, and U. Amann, "Framed: full-fledge role modeling editor (tool demo)," in *Proceedings of the 2016 ACM SIGPLAN International Conference on Software Language Engineering*. ACM, 2016, pp. 132–136.

[95] M. Palyart, I. Ober, D. Lugato, and J.-M. Bruel, "Hpcml: a modeling language dedicated to high-performance scientific computing," in *Proceedings of the 1st International Workshop on Model-Driven Engineering for High Performance and CLoud computing*. ACM, 2012, p. 6.

[96] A. Andersson and J. Krogstie, "Implementation and first evaluation of a molecular modeling language," in *International Conference on Enterprise, Business-Process and Information Systems Modeling*. Springer, 2015, pp. 293–308.

[97] L. Sion, K. Yskout, A. van Den Berghe, R. Scandariato, and W. Joosen, "Masc: modelling architectural security concerns," in *Modeling in Software Engineering (MiSE), 2015 IEEE/ACM 7th International Workshop on*. IEEE, 2015, pp. 36–41.

[98] A. Longo, S. Giacovelli, and M. A. Bochicchio, "Modeling service contracts composition, management and visualization with tree graphs: Ma. vi. c," in *6th International Conference on Management of Emergent Digital EcoSystems*. ACM, 2014, pp. 42–49.

[99] F. B. Aydemir, P. Giorgini, and J. Mylopoulos, "Multi-objective risk analysis with goal models," in *IEEE 10th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2016, pp. 1–10.

[100] B. Roelens, W. Steenacker, and G. Poels, "Realizing strategic fit within the business architecture: the design of a process-goal alignment modeling and analysis technique," *Software & Systems Modeling*, pp. 1–32, 2017.

[101] J. P. Diprose, B. A. MacDonald, and J. G. Hosking, "Ruru: A spatial and interactive visual programming language for novice robot programming," in *Visual Languages and Human-Centric Computing (VL/HCC), 2011 IEEE Symposium on*. IEEE, 2011, pp. 25–32.

[102] M. A. Bochicchio, A. Longo, and S. Giacovelli, "Sara: A tool for service levelsaware contracts," in *Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium on*. IEEE, 2013, pp. 1211–1217.

[103] M. Almorsy and J. Grundy, "Secdsvl: A domain-specific visual language to support enterprise security modelling," in *Software Engineering Conference (ASWEC), 2014 23rd Australian*. IEEE, 2014, pp. 152–161.

[104] B. Henderson-Sellers, G. Low, and C. Gonzalez-Perez, "Semiotic considerations for the design of an agent-oriented modelling language," in *Enterprise, Business-Process and Information Systems Modeling*. Springer, 2012, pp. 422–434.

[105] F. Wanderley, A. Silva, J. a. Araujo, and D. S. Silveira, "Snapmind: A framework to support consistency and validation of model-based requirements in agile development," in *Model-Driven Requirements Engineering Workshop (MoDRE), 2014 IEEE 4th International*. IEEE, 2014, pp. 47–56.

[106] M. Malinova and J. Mendling, "The effect of process map design quality on process management success," in *ECIS*, 2013, p. 160.

[107] C. De Faveri, A. Moreira, J. Arajo, and V. Amaral, "Towards security modeling of e-voting systems," in *Requirements Engineering Conference Workshops (REW), IEEE International*. IEEE, 2016, pp. 145–154.

[108] M. Gómez and J. Cervantes, "User interface transition diagrams for customerdeveloper communication improvement in software development projects," *Journal of Systems and Software*, vol. 86, no. 9, pp. 2394–2410, 2013.

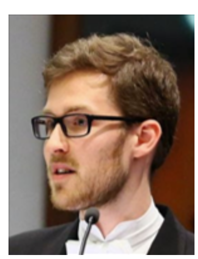
[109] Z. Dobesova, "Visual language for geodatabase design," *13th SGEM GeoConference on Informatics, Geoinformatics And Remote Sensing*, vol. 1, no. International Multidisciplinary Scientific GeoConference SGEM2013, pp. 603–610, 2013.

[110] G. Sindre, J. Krogstie, and S. Gopalakrishnan, *Visually Capturing Usage Context in BPMN by Small Adaptations of Diagram Notation*. Springer, 2013, pp. 324–338.

[111] G. Erdogan and K. Stølen, "Design decisions in the development of a graphical language for risk-driven security testing," in *International Workshop on Risk Assessment and Risk-driven Testing*. Springer, 2016, pp. 99–114.

[112] M. Salnitri, F. Dalpiaz, and P. Giorgino, "Designing secure business processes with secbpmn," *Software & Systems Modeling*, pp. 1–21, 2015.

[113] D. Knuplesch and M. Reichert, "A visual language for modeling multiple perspectives of business process compliance rules," *Software & Systems Modeling*, pp. 1–22, 2016.



Dirk van der Linden is a Senior Research Associate at the Department of Computer Science, University of Bristol. He received the PhD in Information Systems from Radboud University (2015) while working as a researcher at the Luxembourg Institute of Science & Technology. Previously he was an Institutional Postdoctoral Fellow at the Department of Information Systems, University of Haifa. His main research interests center around human aspects of conceptual modeling, requirements engineering, and cybersecurity. Van der Linden has published over 30 papers in international journals and conferences (IJISMD, AO, RE, CBI, CHI PLAY, etc.).

Irit Hadar is a tenured faculty member at the Department of Information Systems, University of Haifa, and the Head of the Software Architecture Laboratory at the Caesarea Rothschild Institute for Interdisciplinary Applications of Computer Science. She received the Ph.D. (2004) in Computer Science Education, from the Technion – Israel Institute of Technology. Her main research area is cognitive aspects of requirements engineering and software architecture and design. Hadar has published over 100 papers in international journals and conferences (CACM, EMSE, JAIS, EJIS, REJ, IST, JSS, RE, OOPSLA, AMCIS, etc.). She has been serving as an organizer and PC member in conferences and workshops (RE, CAiSE, ICIS, AMCIS, MCIS, OOPSLA), and served as an editorial board member of the ACM TOCE journal (2011-2015). Hadar received research grants from the Israel Science Foundation and CA Technologies.

Appendix A
Selected papers

TABLE 2: Selected papers

#	Title	Authors	Source	Type	Publisher	Year	Ref
1	Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation	Moody, Daniel L., Patrick Heymans, and Raimundas Matuleviius	Requirements Engineering	journal	Springer	2010	[40]
2	A model for visual specification of e-contracts	Martinez, Enrique, Gregorio Daz, M. Emilia Cambronero, and Gerardo Schneider	IEEE International Conference on Services Computing (SCC), 2010	conference	IEEE	2010	[54]
3	Cognitive effectiveness of visual instruction design languages	Figl, Kathrin, Michael Derntl, Manuel Caeiro Rodriguez, and Luca Botturi	Journal of Visual Languages & Computing	journal	Elsevier	2010	[55]
4	A taxonomy and visual notation for modeling globally distributed requirements engineering projects	Laurent, Paula, Patrick Mader, Jane Cleland-Huang, and Adam Steele	2010 5th IEEE International Conference on Global Software Engineering (ICGSE)	conference	IEEE	2010	[56]
5	Towards a unified requirements modeling language	Helming, Jonas, Maximilian Koegel, Florian Schneider, Michael Haeger, Christine Kaminski, Bernd Bruegge, and Brian Berenbach	2010 Fifth International Workshop on Requirements Engineering Visualization (REV)	workshop	IEEE	2010	[57]
6	Analysing the Cognitive Effectiveness of the BPMN 2.0 Visual Notation	Genon, Nicolas, Patrick Heymans, and Daniel Amyot	International Conference on Software Language Engineering	conference	Springer	2010	[58]
7	Analysing the Cognitive Effectiveness of the UCM Visual Notation	Genon, Nicolas, Daniel Amyot, and Patrick Heymans	International Workshop on System Analysis and Modeling	workshop	Springer	2010	[59]
8	Using the Physics of notations to analyze a visual representation of business decision modeling	Thomas, John C., Judah Diament, Jacquelyn Martino, and Rachel KE Bellamy	Visual Languages and Human-Centric Computing (VL/HCC), 2012 IEEE Symposium on	conference	IEEE	2012	[41]
9	A visual programming language for designing interactions embedded in Web-based geographic applications	Etcheverry, Patrick, Christophe Marquesuzaà, and Thierry Nodenot	Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces	conference	ACM	2012	[60]
10	MAV-Vis: A Notation for Model Uncertainty	Famelis, Michalis, and Stephanie Santosa	2013 5th Int. Workshop on Modeling in Software Engineering (MiSE)	workshop	IEEE	2013	[61]
11	Evaluation of a graphical modeling language for the specification of manufacturing execution systems	Weißenberger, Benedikt, and Birgit Vogel-Heuser	2012 IEEE 17th Conference on Emerging Technologies & Factory Automation (ETFA)	conference	IEEE	2012	[62]
12	Syntactic and semantic extensions to Secure Tropos to support security risk management	Matulevicius, Raimundas, Haralambos Mouratidis, Nicolas Mayer, Eric Dubois, and Patrick Heymans	Journal of Universal Computer Science	journal	Springer	2012	[63]
13	Vino4TOSCA: A Visual Notation for Application Topologies Based on TOSCA	Breitenbücher, Uwe, Tobias Binz, Oliver Kopp, Frank Leymann, and David Schumm	OTM Confederated International Conferences on the Move to Meaningful Internet Systems	conference	Springer	2012	[64]
14	Using the “Physics” of notation to analyse ModelBuilder diagrams	Dobesova, Zdena	13th Int. Multidisciplinary Scientific GeoConference SGEM 2013	conference	SGEM	2013	[65]
15	A notation for Knowledge-Intensive Processes	Netto, Joanne Manhães, Juliana BS Franca, Fernanda Araujo Baião, and Flávia Maria Santoro	2013 IEEE 17th Int. Conference on Computer Supported Cooperative Work in Design (CSCWD)	conference	IEEE	2013	[46]
16	A visual language for the collaborative visualization of integrated conceptual models in product development scenarios	Herter, Johannes, Ross Brown, and Jivka Ovtcharova	Smart Product Engineering	conference	Springer	2013	[49]
17	Design of visual language syntax for robot programming domain	Plauska, Ignas, and Robertas Damaeviius	Int. Conference on Information and Software Technologies	conference	Springer	2013	[66]
18	Analysing the cognitive effectiveness of the WebML visual notation	Granada, David, Juan Manuel Vara, Marco Brambilla, Verónica Bollati, and Esperanza Marcos	Software & Systems Modeling	journal	Springer	2013	[11]
19	Evaluating and Improving the Visualisation of CHOOSE, an Enterprise Architecture Approach for SMEs	Boone, Sarah, Maxime Bernaert, Ben Roelens, Steven Mertens, and Geert Poels	IFIP Working Conference on the Practice of Enterprise Modeling	conference	Springer	2014	[67]

Selected papers (continued)

#	Title	Authors	Source	Type	Publisher	Year	Ref
20	Towards a more cognitively effective business process notation for requirements engineering	Miske, Carel, Marcus A. Rothenberger, and Ken Peffers	International Conference on Design Science Research in Information Systems	conference	Springer	2014	[68]
21	An evaluation of the statechart diagrams visual syntax	Anwer, Sajid, and Mohamed El-Attar	2014 International Conference on Information Science and Applications (ICISA)	conference	IEEE	2014	[69]
22	Conceptual modeling for ambient assistance	Michael, Judith, and Heinrich C. Mayr	International Conference on Conceptual Modeling (ER)	conference	Springer	2013	[70]
23	A language for process map design	Malinova, Monika	International Conference on Business Process Management	conference	Springer	2014	[71]
24	Using the Physics of Notations Theory to evaluate the visual notation of SEAM	Popescu, George, and Alain Wegmann	2014 IEEE 16th Conference on Business Informatics (CBI). Vol. 2	conference	IEEE	2014	[39]
25	Modelling large-scale information systems using ADLs – An industrial experience report	Woods, Eoin, and Rabih Bashroush	Journal of Systems and Software	journal	Elsevier	2015	[72]
26	A scientific evaluation of the misuse case diagrams visual syntax	Saleh, Faisal, and Mohamed El-Attar	Information and Software Technology	journal	Elsevier	2015	[51]
27	A domain-specific modelling language for clinical pathways in the realm of multi-perspective hospital modelling	Heß, Michael, Monika Kaczmarek, Ulrich Frank, Lars Podleska, and Georg Täger	Twenty-Third European Conference on Information Systems	conference	AIS	2015	[73]
28	GISMO: A domain-specific modelling language for executable prototyping of gestural interaction	Romuald, Deshayes, and Tom Mens	Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems	conference	ACM	2015	[74]
29	The development and experimental evaluation of a focused business model representation	Roelens, Ben, and Geert Poels	Business & Information Systems Engineering	journal	Springer	2015	[75]
30	Jeeves - A visual programming environment for mobile experience sampling	Rough, Daniel, and Aaron Quigley	2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)	conference	IEEE	2015	[76]
31	Towards a context-based representation of the dynamicity perspective in knowledge-intensive processes	Rodrigues, Daya Lages, Flávia Maria Santoro, Fernanda Araujo Baiao, and Joanne Manhães Netto	2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD)	conference	IEEE	2015	[77]
32	A domain-specific visual modeling language for testing environment emulation	Liu, Jian, John Grundy, Iman Avazpour, and Mohamed Abdelrazek	2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)	conference	IEEE	2016	[78]
33	Towards a domain-specific language to design adaptive software: The DMLAS approach	Bocanegram, García, José, Jaime Pavlich-Mariscal, and Angela Carillo-Ramos	Ingeniera y Universidad	journal	Pontificia Universidad Javeriana	2016	[50]
34	A situation-aware workflow modelling extension	Breitenbücher, Uwe, Pascal Hirmer, Kálmán Képes, Oliver Kopp, Frank Leymann, and Matthias Wieland	Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services	conference	ACM	2015	[79]
35	A novel framework for visualizing declarative process models	Hanser, Michael, Claudio Di Ciccio, and Jan Mendling	ZEUS	workshop	CEUR-WS	2016	[80]
36	Empirical validating the cognitive effectiveness of a new feature diagrams visual syntax	Saeed, Mazin, Faisal Saleh, Sadiq Al-Insaif, and Mohamed El-Attar	Information and Software Technology	journal	Elsevier	2016	[81]
37	A visual language for modeling and executing traceability queries	Mäder, Patrick, and Jane Cleland-Huang	Software & Systems Modeling	journal	Springer	2013	[82]
38	A model-based approach for engineering multimodal mobile interactions	Elouali, Nadia, Xavier Le Pallec, José Rouillard, and Jean-Claude Tarby	Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia	conference	ACM	2014	[83]
39	A visual syntax for Larman's operation contracts	Algablan, Abdulaziz S., and Stéphane S. Somé	Engineering & MIS (ICEMIS), International Conference on	conference	IEEE	2016	[84]
40	An explorative analysis of the notational characteristics of the decision model and notation (DMN)	Dangarska, Zhivka, Kathrin Figl, and Jan Mendling	2016 IEEE 20th International Workshop on Enterprise Distributed Object Computing CW	workshop	IEEE	2016	[85]

Selected papers (continued)

#	Title	Authors	Source	Type	Publisher	Year	Ref
41	CAP3: Context-sensitive abstract user interface specification	Van den Bergh, Jan, Kris Luyten, and Karin Coninx	Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems	conference	ACM	2011	[86]
42	Component-based method development: an experience report	Sandkuhl, Kurt, and Hasan Koç	IFIP Working Conference on the Practice of Enterprise Modeling	conference	Springer	2014	[87]
43	Conceptual modeling in human resource management: a design research approach	Strohmeier, Stefan, and Friedrich Röhrs	AIS Transactions on Human-Computer Interaction	journal	AIS	2017	[88]
44	Design of a suite of visual languages for supply chain specification	Zhang, Rick, John Hosking, John Grundy, Nikolay Mehandjiev, and Martin Carpenter	2010 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)	conference	IEEE	2010	[89]
45	Diagram notations for mobile work processes	Gopalakrishnan, Sundar, and Guttorm Sindre	IFIP Working Conference on The Practice of Enterprise Modeling	conference	Springer	2011	[90]
46	Enhancing the communication value of UML models with graphical layers	El Ahmar, Yosser, Sébastien Gérard, Cédric Dumoulin, and Xavier Le Pallec	2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS)	conference	IEEE	2015	[91]
47	Evaluating the appropriateness of the BPMN 2.0 standard for modeling service choreographies: using an extended quality framework	Cortes-Cornax, Mario, Sophie Dupuy-Chessa, Dominique Rieu, and Nadine Mandran	Software & Systems Modeling	journal	Springer	2016	[17]
48	Exploring alternative designs for sociotechnical systems	Aydemir, Fatma Basak, Paolo Giorgini, John Mylopoulos, and Fabiano Dalpiaz	2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)	conference	IEEE	2014	[92]
49	Extending the UML Statecharts Notation to Model Security Aspects	El-Attar, Mohamed, Hamza Luqman, Peter Karpati, Guttorm Sindre, and Andreas L. Opdahl	IEEE Transactions on Software Engineering	journal	IEEE	2015	[93]
50	FRaMED: Full-fledge role modeling editor (Tool demo)	Kühn, Thomas, Kay Bierzynski, Sebastian Richly, and Uwe Aßmann	Proceedings of the 2016 ACM SIGPLAN International Conference on Software Language Engineering	conference	ACM	2016	[94]
51	HPCML: A modeling language dedicated to high-performance scientific computing	Palyart, Marc, Ileana Ober, David Lugato, and Jean-Michel Bruel	Proc. of the 1st Int. Workshop on Model-Driven Engineering for High Performance and Cloud computing	workshop	ACM	2012	[95]
52	Implementation and first evaluation of a molecular modeling language	Andersson, Alexander, and John Krogstie	International Conference on Enterprise, Business-Process and Information Systems Modeling	conference	Springer	2015	[96]
53	MASC: Modelling architectural security concerns	Sion, Laurens, Koen Yskout, Alexander van Den Berghe, Riccardo Scandariato, and Wouter Joosen	7th International Workshop on Modeling in Software Engineering (MiSE), 2015 IEEE/ACM	workshop	IEEE	2015	[97]
54	Modeling service contracts composition, management and visualization with tree graphs: Ma.Vi.C.	Longo, Antonella, Sara Giacovelli, and Mario A. Bochicchio	Proceedings of the 6th International Conference on Management of Emergent Digital EcoSystems	conference	ACM	2014	[98]
55	Multi-objective risk analysis with goal models	Aydemir, Fatma Baak, Paolo Giorgini, and John Mylopoulos	2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)	conference	IEEE	2016	[99]
56	Realizing strategic fit within the business architecture: the design of a Process-Goal Alignment modeling and analysis technique	Roelens, Ben, Wout Steenacker, and Geert Poels	Software & Systems Modeling	journal	Springer	2017	[100]
57	Ruru: A spatial and interactive visual programming language for novice robot programming	Diprose, James P., Bruce A. MacDonald, and John G. Hosking	2011 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)	conference	IEEE	2011	[101]
58	SARA: A tool for service levels-Aware contracts	Bochicchio, Mario Alessandro, Antonella Longo, and Sara Giacovelli	2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013),	conference	IEEE	2013	[102]

Selected papers (continued)

#	Title	Authors	Source	Type	Publisher	Year	Ref
59	SecDSVL: A domain-specific visual language to support enterprise security modelling	Almorsy, Mohamed, and John Grundy	23rd Australian Software Engineering Conference (ASWEC), 2014	conference	IEEE	2014	[103]
60	Semiotic considerations for the design of an agent-oriented modelling language	Henderson-Sellers, Brian, Graham Low, and Cesar Gonzalez-Perez	Enterprise, Business-Process and Information Systems Modeling	conference	Springer	2012	[104]
61	SnapMind: A framework to support consistency and validation of model-based requirements in agile development	Wanderley, Fernando, António Silva, João Araujo, and Denis S. Silveira	2014 IEEE 4th International Model-Driven Requirements Engineering Workshop (MoDRE),	workshop	IEEE	2014	[105]
62	The effect of process map design quality on process management success	Malinova, Monika, and Jan Mendling	Proceedings of the 21st European Conference on Information Systems	conference	AIS	2013	[106]
63	Towards an operationalization of the ”physics of notations” for the analysis of visual languages	Störrle, Harald, and Andrew Fish	International Conference on Model Driven Engineering Languages and Systems	conference	Springer	2013	[24]
64	Towards security modeling of E-voting systems	De Faveri, Cristiano, Ana Moreira, João Araújo, and Vasco Amaral	IEEE International Requirements Engineering Conference Workshops (REW)	workshop	IEEE	2016	[107]
65	User Interface Transition Diagrams for customerdeveloper communication improvement in software development projects	Gómez, M., and J. Cervantes	Journal of Systems and Software	journal	Elsevier	2013	[108]
66	Visual language for geodatabase design	Dobesova, Zdena	13th International Multidisciplinary Scientific GeoConference SGEM 2013	conference	SGEM	2013	[109]
67	Visually capturing usage context in BPMN by small adaptations of diagram notation	Sindre, Guttorm, John Krogstie, and Sundar Gopalakrishnan	Enterprise, Business-Process and Information Systems Modeling	conference	Springer	2013	[110]
68	Design decisions in the development of a graphical language for risk-driven security testing	Erdogan, Gencer, and Ketil Stølen	International Workshop on Risk Assessment and Risk-driven Testing	workshop	Springer	2017	[111]
69	Designing secure business processes with secBPMN	Salnitri, Mattia, Fabiano Dalpiaz, and Paolo Giorgini	Software & Systems Modeling	journal	Springer	2015	[112]
70	A visual language for modeling multiple perspectives of business process compliance rules	Knuplesch, David, and Manfred Reichert	Software & Systems Modeling	journal	Springer	2016	[113]

APPENDIX B
CLASSIFIED RESULTS.

TABLE 3: Classified results. Listed are respectively notation, focus, requirements elicitation, evaluation, justification, and consideration of alternatives. The different properties are denoted by -, ±, and +, respectively denoting the absence of the property, claimed or mentioned, but not shown inclusion of the property, and reporting of the property. In the alternatives column the abbreviations stand for resp. Cognitive Dimensions (CD), SEQUAL Framework, Guidelines of Modeling (GoM), and 7 Process Modeling Guidelines (7PMG).

No.	Notation	Focus	Status	Elicitation of requirements?	Evaluation of design?	Justification of PoN use?	Investigated alternatives?
1	i*	Goal-oriented modeling	Existing	-	-	-	-
2	Contract-Oriented Diagram	Constraints	New	-	±	-	-
3.1	E2ML	Visual instructional design	Existing	±	+	-	-
3.2	PoEML	Visual instructional design	Existing	±	+	-	-
3.3	CoUML	Visual instructional design	Existing	±	+	-	-
4	No name	Distributed requirements engineering processes	New	-	-	-	-
5	URML	Requirements modeling	New	-	-	-	-
6	BPMN 2.0	Processes	Existing	-	-	+	CD, SEQUAL
7	Use Case Map (UCM)	Scenario modeling	Existing	-	-	+	CD, SEQUAL, GoM
8	No name	Business rules & decisions	New	-	-	-	CD
9	No name (a VPL)	Interaction between user and system	New	-	±	-	CD
10	MAV-Vis	Design uncertainty	New	-	±	-	-
11	MES-ML	Requirements engineering	Existing	-	+	-	-
12	Secure Tropos	Security requirements	Existing	-	-	-	-
13	Vino4TOSCA	Topologies	New	+	-	-	-
14	ModelBuilder Diagrams	Dataflow diagrams	Existing	-	-	±	CD
15	KIPN	Knowledge-intensive processes	New	-	-	-	-
16	No name	Model integration	New	-	-	+	CD
17	VisuRobo	Robot programming	New	-	-	-	-
18	WebML	Web and SOA applications	Existing	-	±	+	SEQUAL, GoM, CD
19	CHOOSE	Enterprise Architecture for SMEs	Existing	±	±	+	SEQUAL, CD
20	STSN	Business process modeling	New	-	-	-	-
21	UML Statechart diagrams	Data flow and sequence of events	Existing	-	-	-	-
22	HCM-L	Sequences of daily actions	New	-	-	-	-
23	No name	Process map	New	-	-	-	-
24	Systemic Enterprise Architecture Methodology (SEAM)	Enterprise Architecture	Existing	-	-	+	CD, SEQUAL, 7PMG
25	No name (an ADL)	Software Architecture	New	±	-	-	-
26	Misuse Case Diagrams	Use cases	Existing	-	±	+	CD
27	DSML4CPs	Clinical pathways (hospital processes)	New	+	+	-	-
28	GISMO	Gestural interaction	New	-	±	-	-
29	No name (BM Viewpoint)	Business models	Version	-	±	-	-
30	Jeeves	Visual programming	New	-	-	-	-
31	KIPN-C	Knowledge-intensive processes	Version	-	-	-	-
32	TeeVML	(Software) Testing environment emulation	New	-	-	-	-
33	DMLAS	Design of adaptive systems	New	-	-	-	-
34	SitME	Situational dependencies (in workflows)	Version	-	-	-	-
35	Declare	Declarative process models	New	-	-	±	-
36	Feature diagrams	Feature modeling in software	Existing	-	+	+	CD
37	VTML	Traceability	New	-	+	-	-
38	M4L (Mobile MultiModality Modeling Language)	Multimodal mobile interactions	New	-	±	-	-
39	Visual operation contract	Operation contracts	New	-	-	-	-

Classified results (continued)

No.	Notation	Focus	Status	Elicitation of requirements?	Evaluation of design?	Justification of PoN use?	Investigated alternatives?
40	DMN (Decision Model and Notation)	Decision-making	Existing	-	-	-	-
41	CAP3	User interfaces	New	-	-	-	-
42	CaaS notation	Method development	New	-	+	-	-
43	A CHRM modeling language	Human resources	New	-	-	-	Frank (2013)
44	MaramaSUDDEN	Supply chain specification	New	-	+	-	-
45	UML Activity Diagrams [Adaptation]	Process modeling with location	Version	-	-	-	-
46	FlipLayers	General-purpose	Version	-	-	-	CD
47	BPMN 2.0 Choreographies	Choreographies	Existing	-	-	-	-
48	DESIST	Sociotechnical systems	Version	-	-	-	-
49	UML Statechart diagrams [extension]	Security concerns	Version	-	+	+	CD
50	FRaMED	Roles	New	-	-	-	-
51	HPCML	High-performance computing	New	-	-	-	-
52	GEMAL - Generic Enterprise Modeling and Analysis Language	General-purpose	New	-	+	-	SEQUAL
53	Modelling Architectural Security Concerns (MASC)	Security concerns	Version	-	-	-	-
54	Ma.Vi.C.	Contracts	New	-	±	-	-
55	No name	Risk analysis	New	-	-	-	-
56	Process-goal alignment modeling language	Strategic fit	New	±	-	-	-
57	Ruru	Robot software development	New	-	+	-	-
58	SARA	Contracts	New	-	-	-	-
59	SecDSVL	Enterprise systems security	New	-	+	-	-
60	FAML notation	Agent-oriented IS design	Version	+	+	-	-
61	SnapMind	User-centered requirements modeling	New	+	-	-	-
62	Process maps	Process maps	New	-	-	-	-
63	UML Use Case Diagrams	Use cases	Existing	-	-	+	CD
64	EVSec	Security concerns in voting	New	-	-	-	-
65	User Interface Transition Diagram (UITD)	User interfaces	New	-	±	-	-
66	ARCGis Diagrammer	GIS	Existing	-	-	-	-
67	BPMN [Adaptation]	Process modeling with location	Version	-	-	-	SEQUAL
68	CORAL	Risk assessment of test cases	New	-	-	-	-
69	SecBPMN-ml	Business processes	Version	-	±	-	-
70	eCRG (extended Compliance Rule Graph)	(Business process) Compliance rules	Version	-	±	-	-

Appendix C
Evaluation of detail in principle reporting

TABLE 4
Evaluation of detail in principle reporting. As previously, the degree to which a principle is satisfied is denoted as -, ± and + for resp. no reporting or reasoned absence, ± for claimed satisfaction with no details, + for any kind of details given for claimed satisfaction, and ? for an unverifiable claim.

Paper ID	Semiotic Clarity	Perceptual Discriminability	Semantic Transparency	Complexity Management	Cognitive Integration	Visual Expressiveness	Dual Coding	Graphic Economy	Cognitive Fit
1	+	+	+	+	-	+	-	+	-
2	±	±	?	±	±	?	+	±	?
3.1	+	+	+	?	?	+	+	+	+
3.2	+	+	+	?	?	+	+	+	?
3.3	+	+	+	?	?	+	+	+	?
4	+	+	+	+	+	+	+	+	?
5	±	±	±	±	±	±	±	±	±
6	+	+	+	+	+	+	+	+	+
7	+	+	+	+	+	+	+	+	+
8	+	+	+	-	+	+	+	+	-
9	+	+	+	+	+	+	+	+	+
10	+	+	+	+	+	+	+	+	+
11	+	+	+	+	+	+	+	+	+
12	+	-	-	-	-	-	-	-	-
13	?	+	+	+	+	+	+	+	+
14	+	+	+	-	+	+	+	+	-
15	+	+	-	-	+	-	+	+	-
16	+	-	-	-	-	-	+	-	+
17	?	?	?	?	?	?	?	?	?
18	+	+	+	+	+	+	+	+	+
19	+	+	+	+	-	+	-	-	-
20	?	?	?	?	?	?	?	?	?
21	+	+	?	+	?	+	+	+	?
22	?	?	?	?	?	?	?	?	?
23	?	?	±	±	±	±	±	±	±
24	+	+	+	+	+	+	+	+	+
25	?	?	?	?	?	?	?	?	?
26	+	+	+	+	+	+	+	+	+
27	?	?	?	?	?	?	?	?	?
28	-	±	±	-	-	±	-	-	-
29	+	-	-	+	+	-	-	+	-
30	?	±	?	?	?	?	?	?	?
31	?	?	?	+	?	?	?	?	?
32	+	+	+	+	+	+	+	+	-
33	+	±	±	+	±	+	+	+	+
34	?	?	+	?	?	?	?	?	?
35	+	?	?	+	+	?	?	+	?
36	+	+	+	+	+	+	+	+	+
37	±	±	±	±	±	±	-	±	±
38	+	+	±	+	-	-	+	-	-
39	+	+	+	+	±	+	+	+	+
40	+	+	+	+	+	+	+	?	-
41	?	+	±	-	±	?	?	?	?
42	+	+	+	+	+	+	+	+	-
43	?	+	?	?	?	?	±	?	?
44	?	?	?	±	±	?	±	?	?
45	+	+	+	+	-	+	-	+	-
46	-	±	-	+	+	+	-	-	-
47	+	+	+	+	+	+	+	+	+
48	+	+	?	?	?	+	?	?	?
49	+	+	+	+	+	+	+	+	+
50	?	?	?	?	?	?	?	?	?
51	?	±	±	?	?	?	?	?	?
52	?	?	?	?	?	?	?	?	?
53	+	?	+	+	?	?	+	?	?
54	?	?	?	?	?	?	?	?	?
55	+	+	+	+	?	?	+	±	?
56	?	+	+	?	?	+	?	+	?
57	+	+	+	±	+	+	+	±	?
58	±	+	?	+	+	?	+	+	?
59	+	+	+	+	+	+	?	±	?
60	±	±	±	-	-	±	?	±	±
61	±	±	±	?	?	?	?	?	?
62	+	+	+	+	+	+	+	+	+
63	+	+	-	-	-	-	-	-	-
64	?	?	?	?	?	?	?	?	?
65	+	+	+	+	+	+	+	+	+
66	+	+	+	-	+	+	+	+	+
67	+	+	+	±	-	±	+	±	-
68	+	+	+	+	+	+	+	+	+
69	?	+	?	?	?	?	?	?	?
70	±	±	±	-	-	-	-	±	±