Analysing the Cognitive Effectiveness of the BPMN 2.0 Visual Notation

Nicolas Genon[1], Patrick Heymans[1], and Daniel Amyot[2]

[1] PReCISE, University of Namur, Belgium
{nge,phe}@info.fundp.ac.be
[2] University of Ottawa, Canada
damyot@site.uOttawa.ca

Abstract. BPMN 2.0 is an OMG standard and one of the leading process modelling notations. Although the current language specification recognises the importance of defining a visual notation carefully, it does so by relying on common sense, intuition and emulation of common practices, rather than by adopting a rigorous scientific approach. This results in a number of suboptimal language design decisions that may impede effective model-mediated communication between stakeholders. We demonstrate and illustrate this by looking at BPMN 2.0 through the lens of the Physics of Notations, a collection of evidence-based principles that together form a theory of notation design. This work can be considered a first step towards making BPMN 2.0's visual notation more cognitively effective.

1 Introduction

The Business Process Modeling Notation (BPMN) has recently emerged as the industry standard notation for modelling business processes. Originally developed by the Business Process Management Initiative (BPMI), it is now maintained by the Object Management Group (OMG). It aims to provide a common language for modelling business processes, to replace the multiple competing standards that currently exist. As stated in its latest release (BPMN 2.0 [1][1]), BPMN ambitions to "provide a notation that is readily understandable by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business people who will manage and monitor those processes" [1, p. 28]. Considering the enormous influence that the OMG has in the IT industry, this mission statement holds the promise of delivering a standardised *lingua franca* for all those who, in one way or another, have to deal with business processes.

The standard goes on to say that "[a] key element of BPMN is the choice of shapes and icons used for the graphical elements [. . .]" and that "[the] intent is to create a standard visual language that all process modelers will recognize and understand" [1, p. 29–30]. From these statements, it is clear that a chief

[1] This analysis was performed on version 0.9.15.

B. Malloy, S. Staab, and M. van den Brand (Eds.): SLE 2010, LNCS 6563, pp. 377–396, 2011.

concern of BPMN 2.0 is *cognitive effectiveness* – although the standard does not use that term. Cognitive effectiveness is defined as the *speed*, *ease* and *accuracy* with which a representation can be processed by the human mind [2,3]. A major incentive for using visual notations is the widely-held belief that they convey information more effectively than text, especially to novices [4]. However, cognitive effectiveness is not an intrinsic property of visual notations but something that must be designed into them [3,5].

The proponents of BPMN 2.0 somehow acknowledged this when they state that "[they have] brought forth expertise and experience with many existing notations and [have] sought to consolidate the best ideas from these divergent notations into a single standard notation" [1, p. 28]. However, a striking thing when reading through the voluminous document (504 pages) is the lack of *design rationale*, that is, explicit reference to theories and empirical evidence for designing effective visual notations.

To be perfectly fair to BPMN 2.0, we should recognise that it actually does *better* than many other language definitions, in at least two respects. Firstly, its notation has been defined after reviewing a large number of other notations, including UML Activity Diagrams, UML EDOC Business Processes, IDEF, ebXML BPSS, ADF, RosettaNet, LOVeM, and EPC [1, p. 28]. Secondly, it gives a thorough description of its graphical notation, including an extensive list of the available symbols, how they can be combined and how they can (and cannot) be customised for domain- or application-specific usages.

Still, BPMN 2.0 does not justify its notation design choices by referring to explicit principles. This is actually not surprising since this is a common practice in visual notation design [2]. In this respect, BPMN 2.0 does neither better nor worse than the vast majority of notations: it relies on common sense, intuition and emulation of common practice, rather than adopting a rigorous *scientific* approach.

The objective of this paper is to conduct an analysis of the current BPMN 2.0 visual notation that is based on theory and empirical evidence rather than common sense and opinion. Of course, it is always easy to criticise, but our aim in conducting this analysis is *constructive*: to provide an independent analysis of the strengths and weaknesses of the BPMN visual notation that can be used to improve its usability and effectiveness in practice, especially for communicating with business users. We believe that a unified business process modelling notation is an important innovation and our aim is to help remove potential barriers to its adoption and usage in practice.

A broader goal of this paper is to increase awareness about the importance of visual representation in business process modelling, and the need to refer to theory and empirical evidence in defining notations (evidence-based design practice). Accordingly, our approach enriches emerging research in the area (see Section 2.3). Visual syntax has a profound effect on the effectiveness of modelling notations, equal to (if not greater than) decisions about semantics [6]. But, in contrast to process modelling semantics [7,8], the analysis and definition of process modelling visual notations is a much less mature discipline. Our intention is to remedy this situation.

This paper is structured as follows: Section 2 presents the research background we build upon, and in particular the Physics of Notations, i.e., the theory against which we evaluate BPMN. Section 3 reports on the analysis itself. Section 4 puts the results in a broader perspective and provides a summary of the paper.

2 Previous Research

2.1 Language Evaluation Frameworks

The Cognitive Dimensions of Notations (CDs) framework defines a set of 13 dimensions that provide a vocabulary for describing the structure of cognitive artefacts [9]. This has become the predominant paradigm for analysing visual languages in the IT field. CDs have a level of genericity/generality that makes them applicable to various domains, but it also precludes specific predictions for visual notations. In [9], Green *et al.* noted that the dimensions are vaguely defined and that "the lack of well-defined procedure disturbs some would-be users". These limitations as well as other disadvantages discussed by Moody [10] support our choice not to base our analysis on CDs.

Also popular is the semiotic quality (SEQUAL) framework [11]. It proposes a list of general *qualities* for models and modelling languages, that it organises along the *semiotic ladder* (i.e., the scale 'physical', 'empirical', 'syntactic', 'semantic', 'pragmatic' and 'social'). SEQUAL also distinguishes quality *goals* from the *means* to achieve them, and sees modelling activities as socially situated (*constructivistic worldview*). Essentially, SEQUAL offers a comprehensive ontology of model and modelling language quality concepts. It provides a precise vocabulary and checklist when engaging in a comprehensive analysis. The part of SEQUAL that is most closely related to notation quality is termed 'comprehensibility appropriateness' [12]. However, for our purpose, SEQUAL shares a number of important limitations with CDs (although the two frameworks are very different in intent and content). The two main limitations are the level of generality and the lack of theoretical and empirical foundations related to visual aspects of notations.

2.2 The Physics of Notations

The Physics of Notations theory [2] provides a framework that is *specifically* developed for *visual notations*. It defines a set of 9 evidence-based principles to evaluate and improve the visual notation of modelling languages. The principles are clearly defined and operationalised using evaluation procedures and/or metrics. They are synthesised from theory and empirical evidence stemming from various scientific disciplines such as cognitive and perceptual psychology, cartography, graphic design, human computer interface, linguistics, and communication. This theory is falsifiable [13], i.e., the principles can be used to generate predictions, which are empirically testable. So far, the Physics of Notations has been used to evaluate the visual notations of Archimate [14], UML [15], *i** [16] and UCM [17].

The 9 principles are:

1. *Semiotic Clarity*: there should be a one-to-one correspondence between semantic constructs and graphical symbols.
2. *Perceptual Discriminability*: symbols should be clearly distinguishable.
3. *Visual Expressiveness*: use the full range and capacities of visual variables.
4. *Semantic Transparency*: use symbols whose appearance is evocative.
5. *Complexity Management*: include mechanisms for handling complexity.
6. *Cognitive Integration*: include explicit mechanisms to support integration of information from different diagrams.
7. *Dual Coding*: enrich diagrams with textual descriptions.
8. *Graphic Economy*: keep the number of different graphical symbols cognitively manageable.
9. *Cognitive Fit*: use different visual dialects when required.

Operationalisations of the principles often rely on values of *visual variables*, i.e., the elementary characteristics forming the visual alphabet of diagrammatic notations. The seminal work of Bertin [18] identified 8 visual variables divided into two categories: planar and retinal variables (see Figure 1). Essentially, symbols are obtained by combining visual variable values. Henceforth, we take the convention of underlining visual variable names.

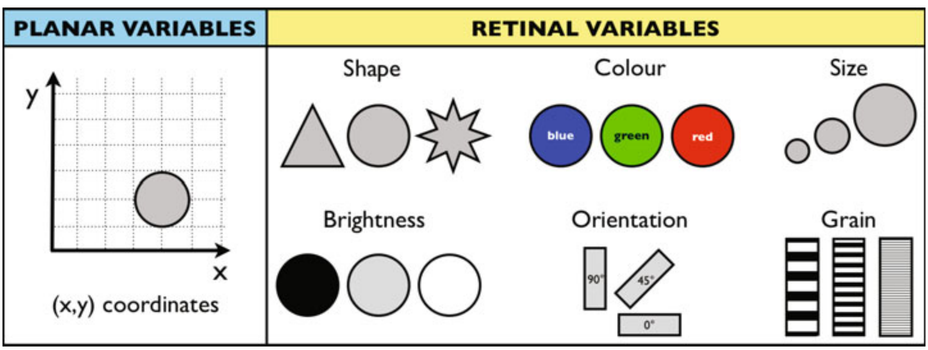

Fig. 1. The 8 visual variables from Bertin [18]

2.3 Visual Aspects of Process Modelling Notations

Studies of the visual syntax of process modelling languages are emerging. Some are concerned with making improvements at the *diagram level* [19,20], whereas our work makes observations and suggestions at the *language level* based on the Physics of Notations. Our work thus focuses on defining notations that are cognitively effective *by construction*. This, of course, does not preclude diagram level improvements (which we actually support) by using so-called *secondary notation* (see Section 3.7). Empirical research has also started to study the impacts of language and context characteristics, such as diagram layout and user expertise [21], routing symbol [22], dual use of icons and labels for process activities [23], ease of use of the language and user experience [24] as well as modularity [25]. Such empirical research is important as it can falsify or corroborate predictions from theories such as the Physics of Notations. However, in the case of BPMN 2.0, we think that application of the theory must come *first* (see Section 4). Nevertheless, we took those studies into account for our analysis of BPMN 2.0, as well work by zur Muehlen and Recker [26] who studied which BPMN concepts are the most frequently used by modellers.

3 Analysis of BPMN 2.0 Process Diagrams

BPMN 2.0 consists of four[2] types of diagrams: PROCESS, CHOREOGRAPHY, COLLABORATION and CONVERSATION diagrams. Process diagrams are by far the most important. The scope of our analysis is limited to them. The 9 principles of the Physics of Notations were thus used to conduct a systematic, symbol-by-symbol analysis of process diagrams. The main findings are reported in the following subsections, organised by principle. For each principle, we provide a definition, summarise the results of the evaluation (how well BPMN satisfies the principle), and give recommendations for improvement. However, we do not go as far as defining a complete new notation which would be overly ambitious and premature. The full analysis can be found in [27].

3.1 Cognitive Fit

As stated by Vessey in the theory of Cognitive Fit [28], there should be a 3-way fit between the *audience* (sender and receiver), the *task characteristics* (how and for what purpose(s) the notation is used) and the *medium* on which the information is represented. BPMN's aim is to "provide a notation that is readily understandable by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business people who will manage and monitor those processes" [1, p. 28]. However, Cognitive Fit theory suggests that trying to design a language that is "all things to all men" is likely to be an impossible mission. Instead, different representations should be used for different tasks, audiences and/or media. BPMN process diagrams can be used in various contexts, e.g., to communicate with non-technical stakeholders, as visual representations of information to be processed by software, etc. The relative importance of the principles differs from one context to another. Different versions (dialects) of a same notation can even be defined for specific contexts. In practice, Cognitive Fit is usually analysed according the two following points of view.

Expert-Novice Differences. There are well-known differences in the way experts and novices create and use diagrams. According to [2], the most important differences are:

[2] The standard does not clearly state if CONVERSATION is a type of diagram or a view on COLLABORATION diagrams.

- Novices have more difficulties discriminating between symbols [4,29].
- Novices have more difficulties remembering what symbols mean [30].
- Novices are more affected by complexity, as they lack mental "chunking" strategies [31].

Concretely, when the purpose of a diagram is to reach a shared high-level understanding of a process among an audience of *novices* (non-technical stakeholders as end users, business experts), then it is particularly important to have a language with few symbols (Graphic Economy) that are mnemonic (Semantic Transparency) and quickly distinguishable (Perceptual Discriminability). Moreover, such diagrams should be kept simple through Complexity Management and Cognitive Integration. On the contrary, if a language is to support detailed discussion between *experts*, or if it is to be fed into a workflow engine, it is more important to have a comprehensive set of symbols with clear semantics (Semiotic Clarity), to be able to represent all required extra details through text (Dual Coding) and to be able to structure large models through Complexity Management and Cognitive Integration. Of course, many other usage contexts could be envisioned, especially when diagrams are devoted to a particular and well-defined task.

In this paper, we chose to illustrate the principles with examples that target an audience of novices. We did so because we think that the BPMN 2.0 notation is more challenging for novices than for experts.

The Differences in Representational Media also support the utilisation of several visual dialects. Rendering diagrams in a computer-based editing tool or drawing them by hand on a whiteboard call for distinct skills. Sketching on a whiteboard requires good drawing abilities, which cannot be assumed from everybody. Indeed, IT practitioners cannot be assumed to be skilled in graphic design. Therefore, sophisticated icons and complex geometrical shapes must be avoided in this kind of notation. Computer-based tools typically do not require such skills (symbols are 'dragged-and-dropped' from a menu onto the diagram) and can easily produce sophisticated visual objects. However, the whiteboard remains a better support for collaborative and interactive modelling [32].

The most common BPMN symbols (see Figure 2A) are basic geometrical shapes. They are easy to draw by hand which is convenient for use on whiteboards, flip charts as well as for paper sketches.

3.2 Semiotic Clarity

According to Goodman's theory of symbols [33], for a notation to satisfy the requirements of a notational system, there should be a one-to-one correspondence between symbols and their referent concepts. Hence, our first task was to inventorise all *semantic constructs* and all visual symbols of process diagrams. At first sight, the list of semantic constructs can be approximated by the list of concrete metaclasses in the BPMN metamodel. However, this is not 100% reliable due to variations in metamodelling styles. A common example is when a construct subtype (e.g., ExclusiveGateway, a subtype of Gateway) can be encoded either as a subclass, or as a particular attribute value of the base metaclass (i.e., GATEWAY). Unless explicitly stated in the standard, this leads to subjective decisions (based on document context and common sense) as to which metamodel elements ought to be considered semantic constructs. We had to make 160 such decisions, adopting consistent policies for similar cases. We eventually counted 242 semantic constructs in process diagrams. Symbols, on the other hand, correspond to legend entries and are thus more straightforward to identify. Process diagrams have 171 symbols, a small sample of which appears in Figure 2. The exhaustive list of all semantic constructs and symbols we took into account is available in [27]. An immediate observation is that these numbers are huge: one or two orders of magnitude beyond languages such as ER (5 symbols), DFD (4 symbols) or YAWL (14 symbols).

Once these numbers are established, Semiotic Clarity can be assessed. Goodman's theory of symbols pinpoints four types of anomalies (analogous to ontological anomalies [7]) that can occur:

- *Symbol deficit*: a construct is not represented by any symbol.
- *Symbol redundancy*: a single construct is represented by multiple symbols.
- *Symbol overload*: a single symbol is used to represent multiple constructs.
- *Symbol excess*: a symbol does not represent any construct.

Semiotic Clarity maximises *expressiveness* (by eliminating symbol deficit), *precision* (by eliminating symbol overload) and *parsimony* (by eliminating symbol redundancy and excess) of visual notations.

We obtained the following results: 23.6% symbol deficit, 5.4% symbol overload, 0.5% symbol excess and 0.5% symbol redundancy. The latter two are negligible. Symbol deficit is the biggest anomaly. It has diverse causes, including domain-specific metaclasses that do not yet have a notation (e.g., AUDITING) or apparent omissions. Symbol overload appears for some constructs, like GATEWAY, which can be represented equally by a diamond or a crossed diamond, with no associated semantic distinction. We see no particular difficulties in removing those anomalies at the notational level, provided that semantic constructs are established. Determining whether modifications to the semantics are required or not is beyond the scope of the Physics of Notations.

3.3 Perceptual Discriminability

Perceptual Discriminability is defined as the ease and accuracy with which different symbols can be differentiated from each other. Discriminability is determined by the visual distance between symbols, which is measured by the number of visual variables on which they differ and the size of these differences (number of perceptible steps between the values of a visual variable). Perceptual Discriminability is a prerequisite for accurate interpretation of diagrams [34]. The greater the visual distance between symbols, the faster and the more accurately they will be recognised [30]. When differences are too subtle, symbol interpretation is much inaccurate. This is especially crucial for novices who have higher requirements for discriminability than experts [4].

Discriminability is a two-step mental process: firstly, symbols are distinguished from the background. This step is called *figure-ground segregation*. Then symbols of different types are discriminated from each other. This second step, called *symbol differentiation*, relies on pair-wise value variations of visual variables between symbols : Shape plays a privileged role in this process, as it represents the primary basis on which we classify objects in the real world [35]. In BPMN process diagrams, 4 shapes are used to derive the majority of symbols. Variations are introduced by changing border style and thickness, and by incorporating additional markers. A selection of these symbols appears in Figure 2A. All shapes are 2-D and from one of two shape families: ellipses (incl. circle) and quadrilaterals (incl. "roundtangle", rectangle and diamond). Ideally, shapes used to represent different semantic concepts should be taken from different shape families to minimise the possibility of confusion [35].

Size is another factor that influences discriminability. Large symbols take more space on the reader's field of vision and focus attention. The size of process diagram symbols is not specified in BPMN 2.0. In practice, the size depends on the amount of text inside the symbol and is thus not related to its semantics. Since Activity appears to be the main symbol type in process diagrams, we would suggest making it explicitly bigger than the 3 other types. Further variations in Size can be used as a secondary notation [21].

Grain (aka. Texture) determines the style and thickness of shape borders. This visual variable can also facilitate symbol differentiation. All BPMN border styles for Events and Activities are shown in Figure 2A. Grain is used to discriminate between 5 types of Events and 4 types of Activities. All 5 visual variable values are distinct, which is a good point, but they quickly become hard to discern when zooming out on the diagram: double lines are merged into a single thick line and dotted/dashed lines become solid lines. Even if this issue is not specific to BPMN, it remains an obstacle to effective discriminability.

Colour is one of the most cognitively effective visual variables: the human visual system is highly sensitive to variations in colours and can quickly and accurately distinguish between them [30]. However, if not used carefully, Colour can undermine communication. In BPMN 2.0, the use of colours is mainly up to tool developers [1, p. 29] (with the exception of the cases discussed in Section 3.7). Nevertheless, the Physics of Notations argues that the choice of colours should be justified by the existence of some sort of association (e.g., logical, metaphorical, rhetorical, cultural) between a symbol and the concept(s) it represents. Additionally, Colour can be used to achieve *redundant coding*. Redundancy is a well-known technique to thwart noise and preserve the signal from errors [36]. Applied to our context, it consists in making symbols distinguishable through concomitant use of several visual variables. Redundant coding is achieved when the value of each visual variable (taken individually) is *unique* among all symbols of the notation.

Based on the above considerations, we illustrate how Perceptual Discriminability can be enhanced in practice, using symbols from BPMN 2.0 (see Figure 3). These new versions of the symbols are motivated by several

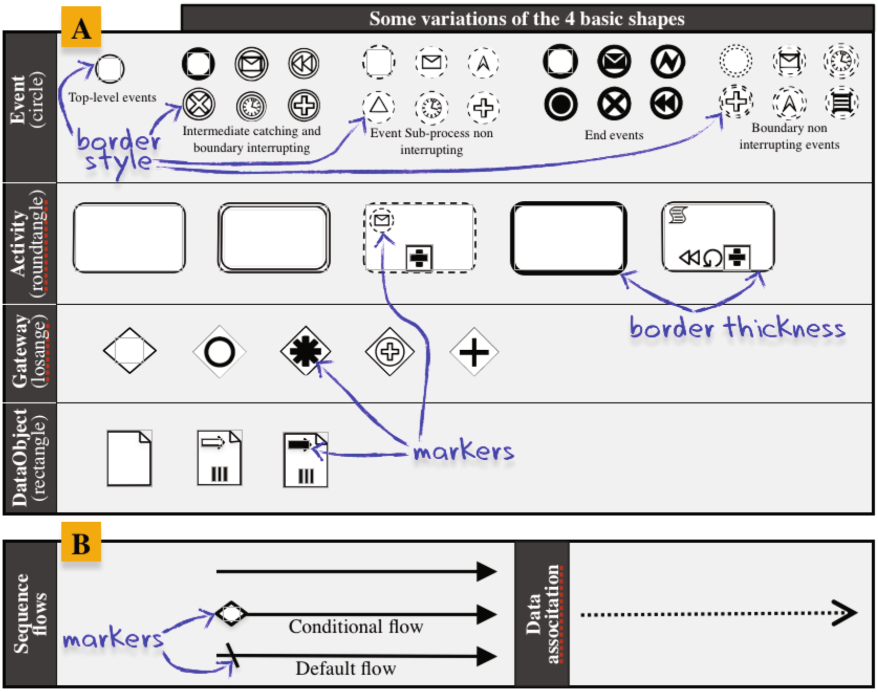

Fig. 2. Basic shapes and connection objects of BPMN 2.0 process diagrams

principles of the Physics of Notations, so further justifications will be provided in the appropriate sections. Here we focus on the rationale related to Perceptual Discriminability:

- We increased discriminability by using 3 shape categories: ellipses, quadrilaterals and cylinders. Further justification of the shapes appears under Semantic Transparency (Section 3.4).
- Border Grain is kept *as is*. It is hard to find another style of border that is easy to draw by hand. However, we propose to improve symbol discriminability by using Colour (see Semantic Transparency).

So far, we have focused on BPMN flow objects, i.e., Event, Activity and Gateway. Relationships between Flow Objects are called Connecting Objects in the BPMN jargon. In BPMN process diagrams, there are two types of relationships: Sequence Flow and Data Association[3]. As shown in Figure 2B, they are represented by monochrome arrows. Sequence Flows can have additional markers at the source, if they are constrained by a condition or if they are the default output of a Gateway. The two representations differ on their Grain: Sequence Flows are solid lines whereas Data Associations are

[3] The BPMN standard defines Message Flow as an element of Collaboration diagrams. Hence they are not considered in this work.

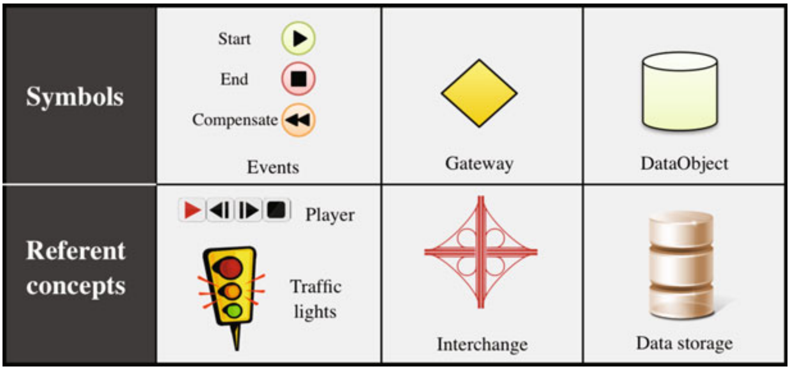

Fig. 3. More semantically transparent symbols proposed for BPMN

dotted. This way, they are distinguishable because they have unique values on at least one visual variable. We should also note that the arrowheads of these two types of flows are distinct (Shape), which reinforces their discriminability. Colour and Grain could also be used to improve their differentiation by introducing redundant coding.

3.4 Semantic Transparency

Symbols should provide cues to their meaning. Semantically direct representations reduce cognitive load through built-in mnemonics: their meaning can be either perceived *directly* or *easily learned* [37]. Such representations speed up recognition and improve intelligibility, especially for novices [4,38]. A symbol's semantic transparency is not a binary state: the transparency level is in a range that goes from semantically *immediate* – the referent concept is understood immediately – to semantically "*perverse*" – the symbol suggests a different meaning.

In BPMN 2.0 process diagrams, symbols are *conventional shapes* on which iconic markers are added (Figure 2A). Symbol shapes seem not to convey any particular semantics: there is no explicit rationale to represent an Event as a circle, an Activity as a roundtangle and a Gateway as a diamond. The situation is even worse for DataObject: its symbol suggests a "sticky note" (a rectangle with a folded corner). This icon is typically used for comments and textual annotations (e.g., in UML), not for first-class constructs. DataObject is thus a case of semantic perversity. The differentiation of Event and Activity subtypes is also purely conventional: it depends on styles of border that are not perceptually immediate.

This lack of semantic immediacy is particularly puzzling, as one of the *leitmotivs* of BPMN is its simplicity of use for novices. Figure 3 proposes more semantically immediate shapes. These shapes are not meant to be the best alternative to the current BPMN 2.0 notation. They are only demonstrations of potential improvements. The suggested Event symbols are inspired from YAWL [8], where they represent types of Conditions using icons inspired from a video player metaphor. Our proposal reuses the BPMN EVENT circle filled with traffic light colours and adorned with player icons. The GATEWAY symbol keeps its diamond shape which suggests the idea of an interchange road sign (yellow diamond). A cylinder has been chosen to represent DATAOBJECT. It stands out from other symbols by being 3-D, thereby also improving discriminability. The cylinder reuses the symbol usually depicting databases or data storage devices, which are in the end closely related.

Regarding Colour, we already justified the choices made for EVENTS. The choice of Colour values should always be based on the nature of the elements represented by the symbols. In practice, this often results from a trade-off that takes into account Symbol and Background Discriminability as well as Semantic Transparency.

Fig. 4. Semantically transparent BPMN icons

The second element that makes BPMN symbols vary is the use of markers. These markers are *icons* (also called mimetic symbols or pictographs) because they perceptually resemble the concepts they represent [39]. Empirical studies show that replacing abstract shapes with icons improves understanding of models by novices [38]. They also improve likeability and accessibility: a visual representation appears more daunting to novices if it is comprised only of abstract symbols [37,40].

BPMN 2.0 includes a large repertoire of icons (more than 25). While the theory recommends *replacing* conventional shapes by icons, BPMN makes a different use of them. They are added *inside* symbols instead of *being the symbols*. Part of the icon repertoire allows distinguishing between subtypes of the 4 basic semantic constructs, while the other part represents attribute values of these constructs. The Semantic Transparency of these icons varies: Figure 4 shows semantically immediate icons that respectively mean: (a) *message*, (b) *manual*, (c) *timer*, (d) *loop* and (e) *exclusive*. If symbols do not appear evocative to novices, once they have been learned, these icons are easily remembered.

On the contrary, Figure 5A illustrates semantically opaque, and in some cases "perverse", icons from BPMN 2.0. The pentagon does not suggest any obvious meaning. In relation to EVENT triggers, it actually means *multiple*. The second icon represents a kind of lightning and could refer to something happening suddenly like an event. In fact, it signifies *error*. The third icon is particularly opaque and even misleading, i.e., it does not mean list but *condition*. The 2 gears, that usually suggest the idea of process or task, are also a case of perversity: this icon refers to the concept of *service* (e.g., web service). The last icon resembles a data sheet, but stands for *business rule*.

In Figure 5B, we suggest new icons that, following the Physics of Notations, are more semantically immediate. As discussed previously, the level of semantic transparency depends on several factors and what seems immediate to someone can remain opaque to somebody else. This suggestion is therefore by no means definitive. The Service icon is a waiter carrying a tray and the Business Rule icon is now composed of a judge's hammer (meaning "rule") filled with a dollar symbol (meaning "business"). Consequently, and even if they may appear semantically perverse at first sight to some people, these icons would probably become immediate for novices as soon as their rationale is *explained*.

Fig. 5. (A) Semantically opaque icons and (B) More semantically transparent icons

Binary directed relationships are classically represented by arrows that are essentially unidimensional (1-D) visual objects. Therefore, only a small portion of the design space is available to achieve Semantic Transparency. Arrowheads are a slight incursion in the 2-D world to show direction. Additionally, the source anchor of SEQUENCE FLOWS can be adorned with a GATEWAY or DEFAULT marker (Figure 2B, left). The GATEWAY marker is rather well chosen whereas the DEFAULT marker is purely conventional. The latter could be removed and the default flow could be indicated by a larger Grain. This technique is used for priority road signs, for example. The representation of DATA ASSOCIATION (see Figure 2B, right) is not semantically immediate but this quality seems hard to achieve for this symbol.

3.5 Complexity Management

One of the major flaws of visual notations is their diagrammatic complexity, which is mainly due to their poor scaling capability [41]. This complexity is measured by the number of elements displayed on a diagram. The degree of complexity management varies according to the ability of a notation to represent information without overloading the human mind. The two main solutions to decrease diagrammatic complexity are *modularisation* and *hierarchic structuring*.

Modularisation consists in dividing complex diagrams into manageable chunks. The decomposition can be *horizontal* or *vertical*. While horizontal decomposition takes place at the same level of abstraction, vertical decomposition produces finer grained sub-diagrams.

BPMN 2.0 provides several mechanisms to manage diagrammatic complexity. First, it supports modelling along 4 different viewpoints that correspond to the 4 types of diagrams: PROCESS, CHOREOGRAPHY, COLLABORATION and CONVERSATION. In a diagram, only the information relevant to the chosen viewpoint has to be represented. BPMN process diagrams achieve modularity thanks to 2 constructs (see Figure 6): (a) LINK EVENTS are used as *intra* or *inter* diagram connectors. They support horizontal decomposition. A LINK EVENT comes as a pair of symbols: a black arrow indicates the source while a white arrow represents the target. Naming the pairs facilitates their association when there are several instances on the same diagram. (b) SUBPROCESSES are self-contained parts of a process. They allow to vertically decompose a diagram in two levelled views: a high-level view – collapsed subprocess – and a fine-grained view – expanded subprocess.

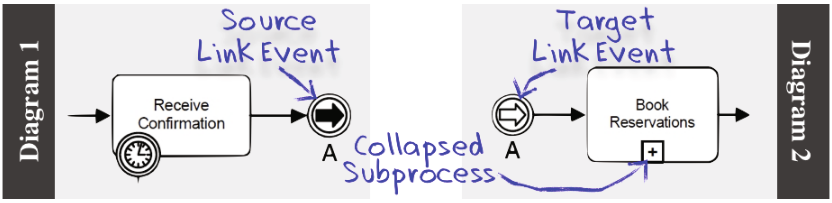

Fig. 6. Modularisation with LINK EVENTS (example from the BPMN 2.0 spec)

Hierarchic Structuring is one of the most effective ways of organising complexity for human comprehension. It allows systems to be represented at different levels of detail, with manageable complexity at each level [42]. SUBPROCESSES provide a means for hierarchic structuring. Yet, to be effective, different levels of information should be displayed in independent diagrams instead of expanding into their parent diagram (see hierarchical *vs.* inline expansion in [43]).

3.6 Cognitive Integration

Large system representations cannot fit into a single diagram. The information is spread across several diagrams and the reader needs to mentally integrate all these pieces of knowledge. Cognitive Integration helps making this integration easier. It takes place at two levels: perceptual integration and conceptual integration. *Perceptual integration* refers to cues that simplify navigation and transitions between diagrams. *Conceptual integration* addresses the assembly of information from separate diagrams into a coherent mental representation.

While Complexity Management leads to the multiplication of diagrams, no technique is available in BPMN to reinforce perceptual or conceptual integration. Mechanisms such as diagram level numbering, signposting and navigation maps [44] could improve perceptual integration. Contextualisation information [45,46] and summary diagrams [47] enhance conceptual integration. Concretely, the notation should ensure that modellers could name their diagrams and number them according to their level in the hierarchic structure. A navigation map could be created based on LINK EVENTS and SUBPROCESSES. Contextualisation is partially achieved as expanded SUBPROCESSES are integrated into their parent ACTIVITY.

3.7 Visual Expressiveness

Visual Expressiveness measures the extent to which the graphic design space is used, i.e., the number of visual variables used by a notation and the range of values for each variable. While Perceptual Discriminability focuses on pairwise visual variation between symbols, Visual Expressiveness measures visual variations across the entire visual vocabulary. Variables that encode information are called information-carrying variables and compose the *primary notation*. The other variables, called the free variables, form the *secondary notation* [21]. They allow modellers to reinforce or clarify the meaning of diagram elements.

Primary Notation. The BPMN process diagram notation uses half of the visual variables: Location (x,y), Shape, Grain and Colour carry semantic information, while Size, Orientation and Brightness are relegated to the secondary notation. Visual variables also have to be chosen according to the type of information to encode. Figure 7 summarises the *power* (highest level of measurement that can be encoded), the *capacity* (number of perceptible steps), the BPMN *values* and the *saturation* (range of values / capacity) of each information-carrying variable.

	Power	Capacity	BPMN Values	Saturation
Location	interval	10 – 15	enclosure	7 – 10%
Shape	nominal	unlimited	circle, roundtangle, diamond, rectangle	-
Grain	nominal	2 – 5	single solid, single thick solid, single dotted, double solid, double dotted	100%
Colour	nominal	7 – 10	black, white	20 – 28%

Fig. 7. Design space covered by the BPMN process diagram notation

We observe that visual variables in BPMN were chosen appropriately according to the nature of information, which here is purely nominal (i.e., there is no ordering between values). Location can actually be used to encode *intervals* but it is used in BPMN only for enclosure (a symbol is contained in another symbol), which is only a small portion of its capacity. Visual variable capacities are rather well exploited and Grain is even completely saturated. However, as we discussed in Section 3.3, this causes discriminability problems. The perceptible steps between Shape values are a major problem of the current notation. Current shapes belong to only two categories (circles and quadrilaterals), whereas there is no semantic relationship between the referent concepts within a shape category. We have already illustrated possible solutions to this problem (see Figure 3).

Colour is one of the most cognitively effective of all visual variables. BPMN 2.0 specification states that "Graphical elements may be colored, and the coloring may have specified semantics that extend the information conveyed by the element as specified in this standard" [1, p. 30]. In fact, BPMN uses only two colours

– black and white – that allow distinguishing between "throwing" (filled) and "catching" (hollow) markers. Hence, the Colour capacity is underused. Improvement proposals have been made for this, e.g., the Event symbols in Figure 3.

Secondary Notation consists of the visual variables that do not carry semantic information, called free variables. At first sight, secondary notation could be thought of as a low priority matter. However, it proves to be of utmost importance when notation engineers face one of the two following situations. The first is when there is a need for introducing a new type of information in the language (e.g., if Semiotic Clarity requires a new symbol to be added in the notation). The language engineer should then first consider free variables before overloading a visual variable that already belongs to the primary notation. This is also the case if the language engineer wants to achieve redundant coding as described in Section 3.3. The second situation occurs when modellers need to place visual cues or hints on the diagram to improve its understandability. Most such cues are *cognitive helpers* that should be implemented with the secondary notation. The main reason is that they do not carry language information, so they should not compete with existing information-carrying variables.

In practice, cognitive helpers are often defined when designing CASE tools. But CASE tool developers usually implement these helpers differently for each tool, resulting in *non standard* solutions even for the same modelling language. To be effective, CASE tools should propose helpers defined by the notation and based on theory and evidence.

3.8 Dual Coding

So far, text has not been considered as an option for encoding information (see Perceptual Discriminability and Visual Expressiveness). However, this does not mean that text has no place in visual notations. According to Dual Coding theory [48], using text and graphics together to convey information is more effective than using either on their own. BPMN makes limited use of Dual Coding. It does so for Conditional and Complex Gateways only. Labels accompany the Alternative or Conditional Flows as appearing in Figure 8. Although we did not observe any major issue with current uses of Dual Coding in BPMN, we suggest to further explore the usage of text in order to improve Graphic Economy, which is BPMN's major problem as discussed in the next section. This contrasts with a recent proposal [49] where Dual Coding is achieved at the expense of Graphic Economy by adding 25 new iconic markers.

3.9 Graphic Economy

Graphic complexity refers to the size of the visual vocabulary, i.e., the number of symbols in a notation [50]. It is measured by the number of legend entries required. This differs from diagrammatic complexity as graphic complexity focuses on the language (type level) rather than the diagram (token level). Graphic Economy seeks to reduce graphic complexity. It is a key factor for cognitive effectiveness since humans' span of absolute judgement when discriminating visual alternatives is around 6 [51]. It can be higher for experts though.

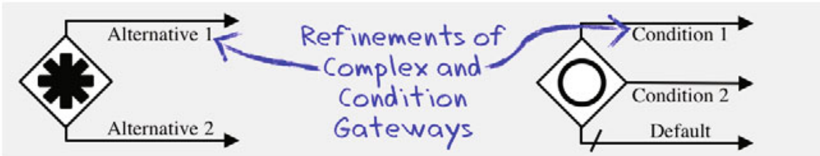

Fig. 8. Dual Coding in BPMN process diagrams

BPMN 2.0 process models have a graphic complexity of 171. This is at least an order of magnitude beyond novice capabilities (28 times the aforementioned limit). zur Muehlen and Recker observe that, in practice, the graphic complexity of BPMN is significantly lower than its nominal complexity [26]. Their study shows that most process diagrams designed for novices use only basic symbols: EVENT, ACTIVITY, GATEWAY, SEQUENCE FLOW, DATAOBJECT and ASSOCIATION, plus a few refinements. The *practical* complexity is thus around 10. This is certainly much more manageable than the full language, but it is still high compared to popular languages [52] such as ER diagrams (complexity of 5) and DFDs (complexity of 4). YAWL, which is more closely related to BPMN, has a complexity of 14. Moreover, the question remains open for BPMN experts: do we really need 171 symbols, even when the goal is to produce detailed models for other experts or for execution in workflow engines? A study similar to that of zur Muehlen and Recker is necessary for such usage contexts too. It could make the case for introducing *symbol deficit* [2], i.e., choosing *not* to provide symbols for some seldom used constructs. Those can still be represented separately using text as suggested in Dual Coding, similar to integrity constraints in ER. It might also be useful to check BPMN for semantic redundancies that could be factored out. Such semantic analyses are beyond the scope of this paper, but at the notation level it is still possible to improve Perceptual Discriminability, Semantic Transparency and Visual Expressiveness as discussed in the previous sections.

4 Discussion and Conclusions

Defining a *cognitively effective* notation is a time-consuming activity governed by conflicting goals. As shown in the analysis presented in this paper, BPMN tries to strike a balance between such goals. But we have argued that it does so in a suboptimal way due to its lack of consideration for existing concepts and scientific principles of notation design. This first complete analysis[4] of BPMN 2.0 against the Physics of Notations theory reveals various problems in the BPMN notation, suggests some improvements, but most importantly recommends a change of methodology.

Given the effort that this would require, we did not go as far as defining a new notation. Our various suggestions are thus not meant to be definitive or consistent with each other. Yet, we deem that they have a value in illustrating

[4] The full analysis is available as a technical report [27].

"outside-of-the-box" thinking about process modelling notations. We hope they will be regarded as sources of *inspiration* and *debate* by the BPM(N) community.

Our analysis (in particular that of Cognitive Fit) also suggests that aiming at a notation that is perfect for all audiences and tasks is utopian. There is no silver bullet. An important question for the future is thus whether there should be one or multiple dialects of BPMN. Silver [43] seems to support this idea with his 3-level methodology for BPM process modelling, using 3 dialects of the BPMN notation. Empirical studies [26] also suggest to restrict the BPMN symbol set when used by novices. But, the *quantity* of symbols (Graphic Complexity) is only one of the dimensions to act upon. For example, in a recent similar analysis of the *i** goal modelling notation [16], Moody *et al.* suggested two dialects, one (for hand sketching) being a *qualitatively* simplified version of the other (for computer-based editing).

An obvious limitation of our research (and the Physics of Notations) is that it focuses only on syntactic issues, whereas solving some of the identified problems (especially the huge number of constructs and symbols) partly requires re-examining the semantics. Another limitation of our work is the lack of empirical validation of *our* suggestions with real BPMN users. This is mitigated by the fact that they are based on theory and empirical evidence synthesised in the Physics of Notations. Moreover, we argue that it would be premature to empirically test these ideas at this stage as they are only *our* suggestions and are not yet fully developed. More work is needed to explore alternative solutions, preferably with participation from BPMN users and researchers. Finally, we acknowledge that there is much legacy related to BPMN as version 1.2 is already used in practice, with support from dozens of commercial tools, some of which cover additional elements from BPMN 2.0. Our contributions may have a limited impact on several legacy symbols but they certainly apply to the numerous new concepts and symbols found in version 2.0, to future versions of BPMN, and to other related languages.

Acknowledgement. The authors are most thankful to Daniel L. Moody for fruitful discussions. This work was supported in part by the Interuniversity Attraction Poles Programme - Belgian State - Belgian Science Policy (MoVES). NSERC provided funding to D. Amyot in support of this collaboration.

References

1. OMG: Business Process Model and Notation (BPMN) Specification 2.0 V0.9.15. Object Management Group Inc. (2009)
2. Moody, D.L.: The "Physics" of Notations: Towards a Scientific Basis for Constructing Visual Notations in Software Engineering. IEEE Transactions on Software Engineering 35, 756–779 (2009)
3. Larkin, J., Simon, H.: Why a Diagram Is (Sometimes) Worth Ten Thousand Words. Cognitive Science 11, 65–99 (1987)
4. Britton, C., Jones, S.: The Untrained Eye: How Languages for Software Specification Support Understanding by Untrained Users. Human Computer Interaction 14, 191–244 (1999)

5. Kosslyn, S.M.: Graphics and Human Information Processing: A Review of Five Books. Journal of the American Statistical Association 80(391), 499–512 (1985)
6. Hitchman, S.: The Details of Conceptual Modelling Notations are Important - A Comparison of Relationship Normative Language. Communications of the AIS 9(10) (2002)
7. Recker, J., Rosemann, M., Indulska, M., Green, P.: Business Process Modeling - a Comparative Analysis. Journal of the Association for Information Systems (JAIS) 10(4), 333–363 (2009)
8. van der Aalst, W., ter Hofstede, A.: YAWL: Yet Another Workflow Language. Information Systems Journal 30(4), 245–275 (2005)
9. Green, T., Blandford, A., Church, L., Roast, C., Clarke, S.: Cognitive Dimensions: Achievements, New Directions, and Open Questions. Journal of Visual Languages and Computing 17, 328–365 (2006)
10. Moody, D.L.: Theory Development in Visual Language Research: Beyond the Cognitive Dimensions of Notations. In: Proc. of the IEEE Symposium on Visual Languages and Human-Centric Computing (VLHCC 2009), pp. 151–154 (2009)
11. Krogstie, J., Sindre, G., Jorgensen, H.: Process Models Representing Knowledge for Action: a Revised Quality Framework. European Journal of Information Systems 15, 91–102 (2006)
12. Krogstie, J., Solvberg, A.: Information systems engineering - conceptual modeling in a quality perspective. In: Kompendiumforlaget, Trondheim, Norway (2003)
13. Popper, K.R.: Science as Falsification. In: Routledge, Keagan, P. (eds.) Conjectures and Refutations, London, pp. 30–39 (1963)
14. Moody, D.L.: Review of archimate: The road to international standardisation. Technical report, Report commissioned by the ArchiMate Foundation and BiZZDesign B.V, Enschede, The Netherlands (2007)
15. Moody, D., van Hillegersberg, J.: Evaluating the visual syntax of UML: An analysis of the cognitive effectiveness of the UML family of diagrams. In: Gašević, D., Lämmel, R., Van Wyk, E. (eds.) SLE 2008. LNCS, vol. 5452, pp. 16–34. Springer, Heidelberg (2009)
16. Moody, D.L., Heymans, P., Matulevičius, R.: Improving the Effectiveness of Visual Representations in Requirements Engineering: An Evaluation of i* Visual Syntax (Best Paper Award). In: Proc. of the 17th IEEE International Requirements Engineering Conference (RE 2009), Washington, DC, USA, pp. 171–180. IEEE Computer Society, Los Alamitos (2009)
17. Genon, N., Amyot, D., Heymans, P.: Analysing the Cognitive Effectiveness of the UCM Visual Notation (to appear, 2010)
18. Bertin, J.: Sémiologie graphique: Les diagrammes - Les réseaux - Les cartes. Gauthier-VillarsMouton & Cie (1983)
19. Mendling, J., Reijers, H., van der Aalst, W.: Seven Process Modelling Guidelines (7PMG). Information and Software Technology 52(2), 127–136 (2010)
20. Mendling, J., Reijers, H., Cardoso, J.: What Makes Process Models Understandable? In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 48–63. Springer, Heidelberg (2007)
21. Schrepfer, M., Wolf, J., Mendling, J., Reijers, H.: The Impact of Secondary Notation on Process Model Understanding. In: Persson, A., Stirna, J. (eds.) PoEM 2009. Lecture Notes in Business Information Processing, vol. 39, pp. 161–175. Springer, Heidelberg (2009)

22. Figl, K., Mendling, J., Strembeck, M., Recker, J.: On the Cognitive Effectiveness of Routing Symbols in Process Modeling Languages. In: Abramowicz, W., Tolksdorf, R. (eds.) BIS 2010. Lecture Notes in Business Information Processing, vol. 47, pp. 230–241. Springer, Heidelberg (2010)
23. Mendling, J., Recker, J., Reijers, H.A.: On the Usage of Labels and Icons in Business Process Modeling. Computer, 40–58 (2010)
24. Figl, K., Mendling, J., Strembeck, M.: Towards a Usability Assessment of Process Modeling Languages. In: Proc. of the 8th Workshop Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (EPK 2009), Berlin, Germany. CEUR Workshop Proceedings, vol. 554, pp. 118–136 (2009)
25. Reijers, H.A., Mendling, J.: Modularity in Process Models: Review and Effects. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 20–35. Springer, Heidelberg (2008)
26. Muehlen, M.z., Recker, J.: How Much Language Is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In: Bellahsène, Z., Léonard, M. (eds.) CAiSE 2008. LNCS, vol. 5074, pp. 465–479. Springer, Heidelberg (2008)
27. Genon, N., Heymans, P., Moody, D.L.: BPMN 2.0 Process Models: Analysis according to the "Physics" of Notations Principles. Technical report, PReCISE - University of Namur (2010), http://www.info.fundp.ac.be/~nge/BPMN/BPMN2_PoN_Analysis.pdf
28. Vessey, I.: Cognitive Fit: A Theory-based Analysis of the Graphs versus Tables Literature. Decision Sciences 22, 219–240 (1991)
29. Koedinger, K., Anderson, J.: Abstract planning and conceptual chunks: Elements of expertise in geometry. Cognitive Science 14, 511–550 (1990)
30. Winn, W.: An Account of How Readers Search for Information in Diagrams. Contemporary Educational Psychology 18, 162–185 (1993)
31. Blankenship, J., Danseraeau, D.F.: The effect of animated node-link displays on information recall. The Journal of Experimental Education 68(4), 293–308 (2000)
32. Persson, A.: Enterprise Modelling in Practice: Situational Factors and their Influence on Adopting a Participative Approach. PhD thesis, Department of Computer and Systems Sciences, Stockholm University (2001)
33. Goodman, N.: Languages of Art: An Approach to a Theory of Symbols. Bobbs-Merrill Co., Indianapolis (1968)
34. Winn, W.: Encoding and Retrieval of Information in Maps and Diagrams. IEEE Transactions on Professional Communication 33(3), 103–107 (1990)
35. Biederman, I.: Recognition-by-Components: A Theory of Human Image Understanding. Psychological Review 94(2), 115–147 (1987)
36. Green, D.M., Swets, J.A.: Signal Detection Theory and Psychophysics. Wiley, Chichester (1966)
37. Petre, M.: Why Looking Isn't Always Seeing: Readership Skills and Graphical Programming. Communications of ACM 38(6), 33–44 (1995)
38. Masri, K., Parker, D., Gemino, A.: Using Iconic Graphics in Entity Relationship Diagrams: The Impact on Understanding. Journal of Database Management 19(3), 22–41 (2008)
39. Pierce, C.: The Essential Writings. Prometheus Books (1998)
40. Bar, M., Neta, M.: Humans Prefer Curved Visual Object. Psychological Science 17(8), 645–648 (2006)
41. Citrin, W.: Strategic Directions in Visual Languages Research. ACM Computing Surveys 24(4) (1996)
42. Flood, R., Carson, E.: Dealing with Complexity: an Introduction to the Theory and Application of Systems Science. Plenum Press, New York (1993)

43. Silver, B.: BPMN Method and Style. Cody-Cassidy Press (June 2009)
44. Lynch, K.: The Image of the City. MIT Press, Cambridge (1960)
45. Lamping, J., Rao, R.: The Hyperbolic Browser: a Focus + Context Technique for Visualizing Large Hierarchies. Journal of Visual Languages and Computing 7, 33–55 (1999)
46. Turetken, O., Schuff, D., Sharda, R., Ow, T.: Supporting Systems Analysis and Design Through Fisheye Views. Communications of ACM 47(9), 72–77 (2004)
47. Kim, J., Hahn, J., Hahn, H.: How Do We Understand a Systeme with (So) Many Diagrams? Cognitive Integration Processes in Diagrammatic Reasoning. Information Systems Research 11(3), 284–303 (2000)
48. Paivio, A.: Mental Representations: A Dual Coding Approach. Oxford University Press, Oxford (1986)
49. Mendling, J., Recker, J., Reijers, H.: On the Usage of Labels and Icons in Business Process Modeling. International Journal of Information System Modeling and Design, IJISMD (2009)
50. Nordbotten, J., Crosby, M.: The Effect of Graphic Style on Data Model Interpretation. Information Systems Journal 9(2), 139–156 (1999)
51. Miller, G.A.: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. Psycological Review, 81–97 (1956)
52. Davies, I., Green, P., Rosemann, M., Indulska, M., Gallo, S.: How Do Practitioners Use Conceptual Modelling in Practice? Data and Knowledge Engineering 58, 358–380 (2006)