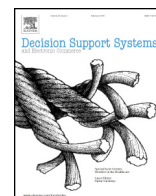




Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: [www.elsevier.com/locate/dss](http://www.elsevier.com/locate/dss)

# How visual cognition influences process model comprehension

Razvan Petrusel<sup>a,\*</sup>, Jan Mendling<sup>b</sup>, Hajo A. Reijers<sup>c</sup>

<sup>a</sup> Universitatea Babeș-Bolyai, Business Informatics Research Center, Strada Teodor Mihali Nr. 58-60, 400591 Cluj-Napoca, Romania

<sup>b</sup> Wirtschaftsuniversität Wien, Welthandelsplatz 1, 1020 Vienna, Austria

<sup>c</sup> Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081HV Amsterdam, The Netherlands

## ARTICLE INFO

### Article history:

Received 30 March 2016

Received in revised form 17 January 2017

Accepted 17 January 2017

Available online xxxx

### Keywords:

Business process model comprehension  
Visual cognition factors of model comprehension  
Theoretical perspective model comprehension  
Eye-tracking experiment business process model  
Complexity of process model comprehension task

## ABSTRACT

Process analysts and other professionals extensively use process models to analyze business processes and identify performance improvement opportunities. Therefore, it is important that such models can be easily and properly understood. Previous research has mainly focused on two types of factors that are important in this context: (i) properties of the model itself, and (ii) properties of the model reader. The work in this paper aims at determining how the performance of subjects varies across *different types of comprehension tasks*, which is a new angle. To reason about the complexity of comprehension tasks we take a theoretical perspective that is grounded in visual cognition. We test our hypotheses using a free-simulation experiment that incorporates eye-tracking technology. We find that model-related and person-related factors are fully mediated by variables of visual cognition. Moreover, in comparison, visual cognition variables provide a significantly higher explanatory power for the duration and efficiency of comprehension tasks. These insights shed a new perspective on what influences sense-making of process models, shifting the attention from model and reader characteristics to the complexity of the problem-solving task at hand. Our work opens the way to investigate and develop effective strategies to support readers of process models, for example through the context-sensitive use of visual cues.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Business process models play an important role in different phases of the business process management lifecycle [1]: These models structure the overall process landscape, they serve as input for analysis, and they can be used as blueprints for process implementation. Business process models (or *process models* for short) are created and utilized collaboratively by process analysts, process owners, process participants, and senior management. They should be presented and designed in such a way that these different stakeholders can best utilize them for the respective tasks at hand.

A prerequisite for an effective usage of process models is that stakeholders can readily understand them. Recent research has investigated process model comprehension by evaluating different types of factors, including model complexity [2,3] as well as model reader characteristics [4–6]. What if we now consider the same model and the same model reader while the comprehension tasks differ? Existing work does not provide any explanation why certain comprehension tasks appear to be easy to solve and others difficult [7]. Yet, understanding the reasons

why certain comprehension tasks are difficult bears the potential to support modeling in a more effective way. First of all, based on such insights, tool features can be designed to help the model viewer in reading and understanding a model. Second, modelers can be directed to those parts of their model that are likely to be difficult to understand by the intended readership.

In this paper, we address this research gap from a theoretical angle. We analyze the comprehension process from the perspective of visual cognition in order to build hypotheses of comprehension task performance in relation to process models. We test our hypotheses using a free-simulation experimental design [8] in order to integrate visual cognition data from an eye-tracking device. The results underline the importance of visual cognition for process model comprehension. Factors associated with visual cognition explain a good share of the overall variance in comprehension performance and mediate classical factors such as model complexity and personal differences. This has implications for designing process models in practice and for research on conceptual models altogether.

The rest of the paper is structured as follows. Section 2 summarizes prior research on process model comprehension and develops hypotheses based on visual cognition. Section 3 presents the design of our study, and Section 4 provides the results. Section 5 discusses implications of this research. Section 6 concludes the paper and points to directions of future research.

\* Corresponding author.

E-mail addresses: [razvan.petrusel@econ.ubbcluj.ro](mailto:razvan.petrusel@econ.ubbcluj.ro) (R. Petrusel), [jan.mendling@wu.ac.at](mailto:jan.mendling@wu.ac.at) (J. Mendling), [h.a.reijers@vu.nl](mailto:h.a.reijers@vu.nl) (H.A. Reijers).

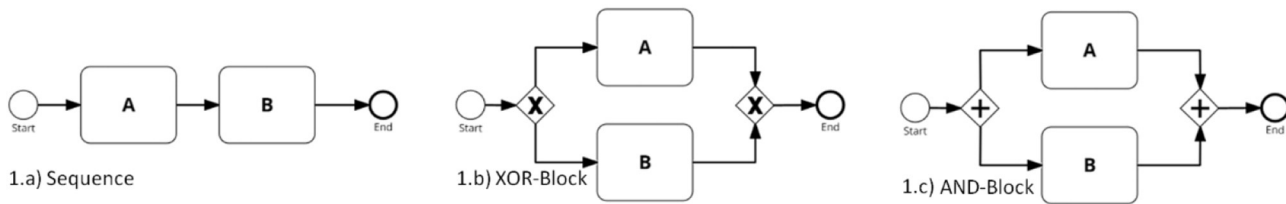


Fig. 1. Control flow representation in BPMN process models.

## 2. Background

In this section, we present the background of our research. First, we summarize prior research on process model comprehension. Then, we discuss visual cognition and its link to the notion of a relevant region. Finally, we present our research question along with corresponding hypotheses.

### 2.1. Process model comprehension

Processes are typically modeled using graphical languages, for example the Business Process Model and Notation (BPMN) (<http://www.omg.org/spec/BPMN/2.0/>). Fig. 1 shows some of the essential building blocks of a BPMN model. The simplest control flow that can be expressed is a sequence (Fig. 1.a). This represents that once activity A is completed, activity B can start. Fig. 1.b) shows a decision, as an XOR-block, modeling a choice represented by a so-called gateway (diamond shape with an x): the process has to continue either with A or B, but not both. Fig. 1.c) illustrates how concurrency is modeled. The AND-split (diamond shape with a +) triggers both branches such that A and B can be executed independently from one another, while any next step to be executed is feasible only after both A and B are finished.

In practice, process models are complex and often deviate from the simple block structures shown in Fig. 1. For example, Fig. 2 shows a BPMN process model from [1] that is already somewhat harder to understand. This model shows that a process can be triggered by an *Incoming call*. The first task, *Call registration*, leads to three possible courses of action following the XOR-split. If for example, an *External Referral with form B4* is received, two tasks need to be executed following the AND-split. Only after both the tasks *Telephone confirmation to external party* and *Archiving system* are completed, synchronized by an AND-join, the *Inform complainant* task can be executed, which completes the process. Typical comprehension tasks for such a BPMN model are questions like “Are the tasks *Telephone confirmation to external party* and *Incident agenda* exclusive to one another?” or “Is *Archiving system* always the last step to perform in the process?”

The prerequisite for a process model to be useful is that it can be readily understood by the involved stakeholders. Research into process model comprehension is, therefore, concerned with identifying measures that capture comprehension effectiveness and efficiency, as well as the factors that make comprehension easy or difficult [9]. Comprehension in this context is measured using comprehension questions as tasks, which help to assess if a person can correctly determine the behavioral relationship between activities in a process model (e.g. concurrency, exclusiveness, sequence, etc.) [2,10]. The performance of answering such tasks in terms of accuracy (i.e. giving the correct answer to a comprehension question) and duration (i.e. how fast the answer is given) can then be used to measure comprehension [11]. Factors that have an impact on comprehension include model characteristics, language characteristics, and personal characteristics.

*Model characteristics* include the size as the number of model elements and complexity as the number of connections between these elements: the bigger and the more complex the model, the more difficult it has been found to be understood [12]. For example, the model in Fig. 2 is difficult to comprehend because its structure that involves six gateways is complex. One example of complex gateway behavior is the AND-split and XOR-join combination that links the two exclusive branches *External referral with form B4* and *Internal referral with form B2*. Various ways to operationalize size and complexity have been used yielding comparable results [3,13,14]. Most prominently, *structuredness* appears to be of specific relevance in this context [15]. The model in Fig. 2 is not structured, since there are split gateways that do not directly match a corresponding join gateway of the same type. For example, one would expect the XOR-join before the *Archiving system* task in the middle of the model and the XOR-join before the last AND-join of the model, to have a corresponding XOR-split.

*Modeling language* that have an impact on comprehension can be related to, first, the formal concepts covered and, second, the notational symbols. Deficiencies in both these matters tend to affect comprehension negatively [16,17]. Also, language complexity seems to be an issue that modelers often try to sooth by restricting the symbol set [18].

Finally, *personal characteristics* have been found to be important for comprehension. Performance of experts appears to be much better

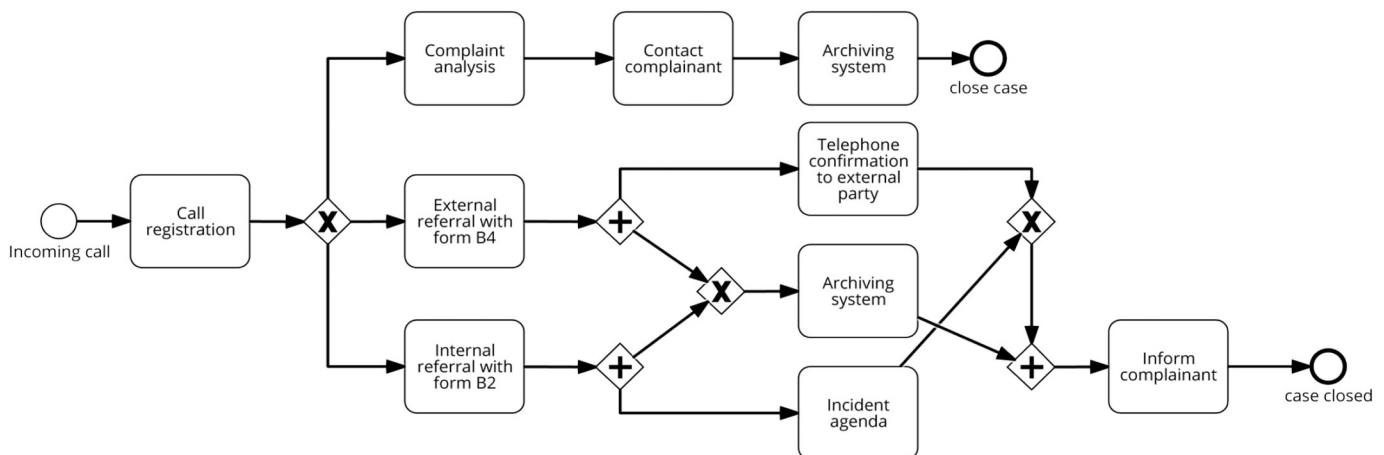


Fig. 2. Example of a BPMN process model for complaint handling with quality issues [1].

than that of novices [2,4]. Different metrics used to capture model reader characteristics include experience with BPMN modeling [19], performance in modeling courses, scores in knowledge tests [3], and familiarity with modeling [11].

Research into other factors is scarce. The work presented in [20] demonstrates the benefits for comprehension of coloring corresponding elements. Similar benefits of coloring have been studied using eye-tracking [21]. The way how activity labels are formulated is relevant as well [22]: the verb-object style as used in *Inform Complainant* is recommended, while deviations such as *External referral with form B4* in Fig. 2 should be avoided. Finally, recent research exists that analyzes in how far certain types of comprehension tasks might be easier or more difficult [7]. This perspective of using tasks as the unit of analysis is novel in research on factors of model comprehension. It is important since it bears the potential to explain differences of performance between different problem-solving tasks that are performed by the same person on the same model. With our research, we follow up on the mentioned study.

## 2.2. Process models and visual cognition

In order to better understand why certain comprehension tasks are difficult, we have to analyze how a model reader approaches them. Typical comprehension tasks ask for the behavioral relationship between two activities of a process model. The model reader has to step through the model to clarify this relationship. An effective strategy for doing so is to analyze the path to the first and to the second activity from the start, and identify the point where both paths diverge. If this point of diversion is, for example, an XOR-split gateway, both activities are likely to be exclusive to one another.

Medical and psychological research describes the so-called eye-mind relationship, which means that we can accurately perceive something only if we fixate it with our eyes and focus our minds on it (i.e. this is commonly named *attention*) [23,24]. The focus of attention from one model element to the other can be captured by an outside observer as eyes fixating the different model elements. Presumably, model readers can solve a comprehension task accurately if they focus their attention on the part of the model that is specifically relevant for it. This idea was first formulated for process models by Petrusel and Mendling [25] with the definition of a so-called *Relevant Region (RR)*. The Relevant Region is the sub-set of process model elements that need to be inspected for solving the comprehension task. Consider Fig. 2 again. If we are

interested in the behavioral relationship of, for example, *Telephone confirmation to external party (TC)* and *Incident Agenda (IA)*, we have to inspect all gateways from the start to both these activities (i.e., one XOR-split and two AND-splits). The first XOR-split from the left is the diverging point of both paths. Therefore, we know that the decision taken here determines whether *TC* or *IA* is executed. This means that both these activities are exclusive to one another. The rest of the model is irrelevant for solving this task. There is a formal definition of the Relevant Region based on graph-theoretic concepts such that it can be automatically calculated with software [21]. The Appendix A shows examples of Relevant Regions highlighted in process models.

The advantage of the Relevant Region is that it can be empirically investigated without asking persons what they think. Given the eye-mind relationship, it links *attention* to an *area of interest*. Eye-tracking can be used to identify this focus of attention by measuring where someone's eyes are fixated. The eye-tracking observation method records several metrics that are interesting in this context [26,27]: the number of fixations (pause of eye movements on a specific area of the visual field), the duration of each fixation, the saccades (rapid movement between two fixations), and the sequence of fixations. Fig. 3 shows eye-tracking output on those metrics. We collectively refer to this group as factors that relate to *Visual Cognition Intensity*. The fixations are displayed as circles, while the sequence and the fixation duration are given by the numbers inside the circles. For answering the comprehension task at the bottom, the subject fixated three elements of the model (see the red, blue, and yellow circles) of which one gateway got the most attention, being fixated twice. These elements are also the crucial gateways of the Relevant Region, highlighted with squares in Fig. 3. Our try-outs strongly suggest the connection between the Relevant Region and the comprehension task [25].

The different types of eye-tracking related data allow us to define measures of how well a model reader is inspecting a model. We refer to such aspects as *Visual Cognition Efficiency*. We approach this measurement from the perspective of *information retrieval*. The notion of relevance is a fundamental concept in that area of research, providing the basis for the two key retrieval metrics: precision and recall [28]. Both describe how well a search system is able to answer queries. Precision is defined as the share of relevant documents in relation to a total of all documents that are returned for a query. Recall describes the share of relevant documents in relation to all relevant documents that should have been retrieved.

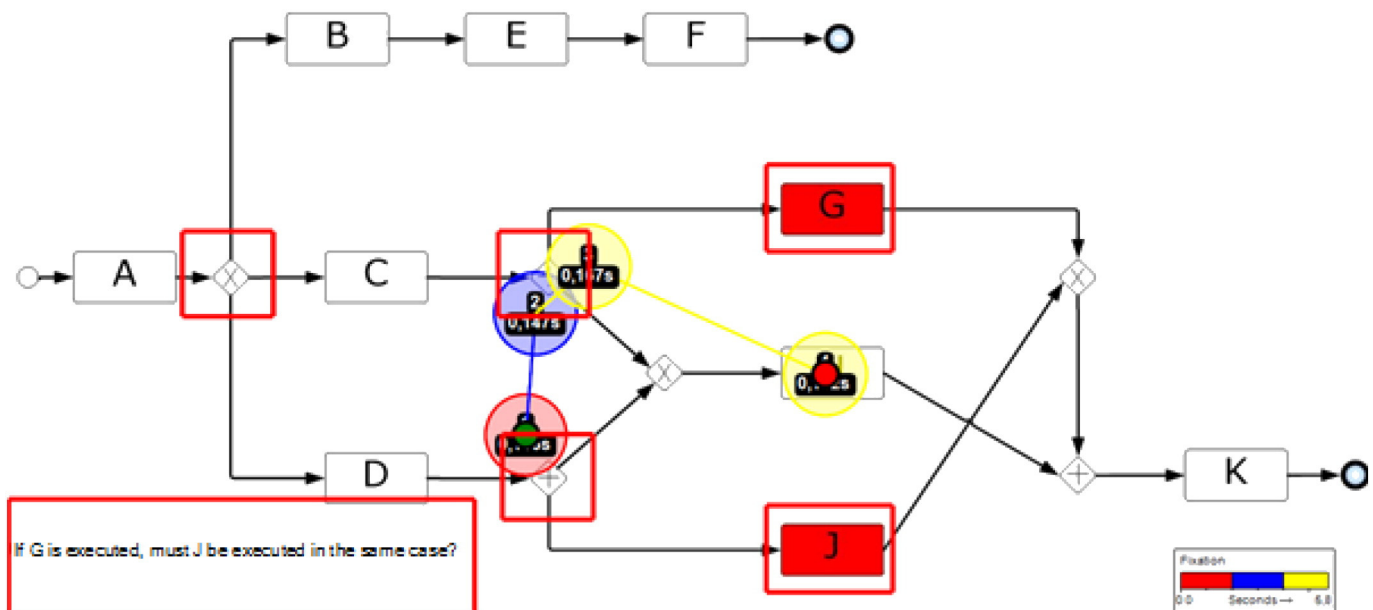


Fig. 3. Eye-tracking output on fixations for the complaint handling process model.

Precision and recall can be directly applied to our problem of visual problem solving. The problem-solving task can be regarded as a query, while the search system is the model reader inspecting the model. The notion of a Relevant Region divides the model into two areas: the relevant model elements and the irrelevant model elements for a specific task. We can then define the two measures based on the search path. We define *Scan-Path Precision* as the percentage of fixations on relevant elements in relation to all fixations. For example, in Fig. 3, this is 75% as three of four fixations are on Relevant Region elements. *Scan-Path Recall* is defined as the percentage of relevant elements that were fixated divided by the number of elements that are relevant for the particular task. For Fig. 3 this metric is 67% as only two out of three relevant gateways were fixated.

### 2.3. Hypotheses

Previous research on process models links Model Complexity to measures such as the number of model elements [3]. Researchers agree that the more complex a model, considering any metric, the more difficult it is to understand. This argument is grounded in Cognitive load theory, which assumes a limited processing capacity of working memory [29]. As eyes move around, perceptual continuity is maintained by placing an encoded representation of the stimuli in visual working memory (VWM). It was shown that VWM is limited to four items at one time [30]. Humans primarily cope with VWM limitation by filtering and processing only the most relevant incoming information. A second successful strategy [31] is to group together objects into sets. Applying those strategies is a matter of training. A single-feature object or an integrated set of objects is processed in just the same way [30]. Therefore, it was established that expertise plays a major role in dealing with complexity. It was also observed that experts perform better than novices in process model comprehension tasks [3,20,10]. Experts are able to build chunks of information, like in our case XOR-blocks, as if they were single elements such that working memory load is reduced [29]. State-of-the-art research on process model reading and understanding describes performance as being influenced by both Model Complexity and Personal Knowledge e.g. [1,13–15,20]. One can rule out factors such as object complexity given that modeling languages use simple graphical representations (e.g. rectangles, arrows) [32]. Assuming a minimal knowledge of the modeling language, the effort to decode model element significance can also be discarded. Therefore, perceptual complexity and, ultimately, comprehension performance are linked to how relevant elements are identified and manipulated and, subsequently, how VWM is filled with objects.

The explanatory model proposed in this paper argues that process model comprehension is based *directly* on the ability of the reader to identify the limited number of model elements relevant for the concrete task to be executed. In this way, the effects of the two previous dimensions are mediated by a third one: the cognitive process of the model reader (Fig. 4). This cognitive process has essentially two facets: (1) the Visual Cognition Intensity with which a subject inspects a model and (2) the Visual Cognition Efficiency (see Fig. 4). It is an open research question to what extent both facets of Visual Cognition mediate the relationship between Model Complexity and Personal Knowledge on the one hand and Comprehension Performance on the other.

Building on Cognitive Load Theory, we formulate the following hypotheses:

**H1.** Higher Personal Knowledge and lower Model Complexity lead to better Visual Cognition Efficiency (measured by Total Fixations and Total Duration of Fixations).

**H2.** Lower Personal Knowledge and higher Model Complexity lead to higher Visual Cognition Intensity (measured by Scan Path Precision and Recall).

**H3.** A Visual Cognition Efficiency (measured by Scan Path Precision and Recall) and Visual Cognition Intensity (measured by Total Fixations and Total Duration of Fixations) model better explains comprehension performance (higher Correctness, higher Efficiency, lower Duration) than a Personal Knowledge and Model Complexity model.

**H4.** The effect of Personal Knowledge on comprehension performance is mediated by Visual Cognition Efficiency and Intensity.

**H5.** The effect of Model Complexity on comprehension performance is mediated by Visual Cognition Efficiency and Intensity.

### 3. Research method

To examine the role of visual cognition in process model comprehension, we designed a free-simulation experiment [8] based on an eye-tracking observation method. Free-simulation experiments are different from traditional factorial experiment designs in that subjects are confronted with tasks and asked to respond to them. Therefore, we can use them to uncover connections between variables that are not exactly binary factors. Our research is specifically focused on visual cognition efficiency and intensity, but these cannot be directly influenced. Free-simulation experiments aim to generate data that shows sufficient variation for each of these metric-scale factors. A consequence of such a research design is that there is no explicit control versus treatment comparison. The data generated from a free-simulation experiment is analyzed with correlation and regression methods.

#### 3.1. Participants

The experimental subjects were 75 experienced modelers from industry and academia. Experiments were conducted with professionals (44% of subjects from companies such as Camunda Services GmbH, Berlin, Germany; Perceptive Software, Apeldoorn, The Netherlands; Signavio GmbH, Berlin, Germany) and academics (56% of subjects from universities such as HPI Potsdam, HU Berlin, TU Eindhoven, UBB Cluj-Napoca, WU Vienna) in cities across Europe (Vienna, Berlin, Potsdam, Apeldoorn, Eindhoven, Cluj-Napoca, and St. Gallen). All subjects participated voluntarily and no reward was provided.

The personal knowledge of the participants was assessed using the three metrics of familiarity in Table 1. The participants have an above-average knowledge of BPMN (both on general knowledge, and on reading BPMN models) and have read on average about 40 models in the last year (some outliers contribute to high variance). With these statistics, we have evidence that the participants were indeed experienced. The number of participants included in the analysis is 72 because we discarded all observations linked to unreliable eye-tracking data as well as one outlier in terms of duration.

#### 3.2. Experimental procedure

The experiment was conducted in four phases: a) Demonstration phase: a question based on a sample model along with experiment and eye-tracking tool description; b) Task Block 1: calibration followed

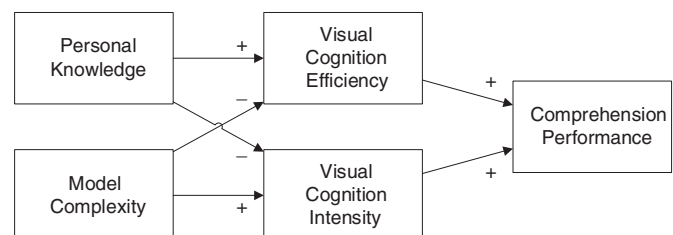


Fig. 4. Visual cognition model of process model comprehension performance.



**Table 1**  
Familiarity metrics of experiment participants.

Variable name	Sample	Min	Max	Mean	Std. Dev.
Familiarity with BPMN	72	1	5	3.43	1.173
Familiarity with reading BPMN models	72	1	5	3.57	1.046
Number of any kind of process models read in the last year	72	0	200	38.50	41.702

by a set of 12 questions, out of which observations from the first 8 were retained for further analysis; c) Task Block 2: re-calibration followed by a second set of 12 questions, out of which observations from the first 4 were retained for further analysis; d) Follow-up: while still using the software tool, participants answered self-assessment familiarity questions, as well as theoretical questions related to the topics under investigation. As we conducted a larger experiment aimed at testing several hypotheses, data collected in connection with 12 questions was used for further analysis in this paper.

This experimental procedure is supported by the S2 eye-tracking system provided by Mirametrix ([www.mirametrix.com](http://www.mirametrix.com)). The hardware includes a binocular, video-based remote eye-tracker, which uses cameras to capture eye movements. As tracking parameters, we used “bright pupil”, a screen resolution of 1280 \* 1024 on a 19 in. screen, and a tracking rate of 60 Hz with an average accuracy between 0.5 and 1 degree, which translates to an error between 15 and 30 pixels. For the fixation duration threshold, we used the default setting of 0.1 s.

The experimental software consists of three tools: EyeMetrix Design, EyeMetrix Record, and EyeMetrix Analyze. The EyeMetrix Design tool allowed us to design the experiment in a slide-show manner. A group of 3 slides were used to operationalize each comprehension question: Slide 1 shows the comprehension question; Slide 2 shows the model (question was repeated at the bottom); Slide 3 asks the subject to provide the answer (by clicking one of the two Yes/No radio buttons).

### 3.3. Experimental tasks

For the experiment, we designed twelve BPMN process models with a comprehension task for each of these. First, we developed our material based on process models that we have encountered in practice or in other publications. For instance, Model 6 is based on the model shown as Fig. 2 in the background section, which stems from [1]. We chose models that are not simple, but also not overly complex. The display size of the computer screen in the eye-tracker provided an upper bound. We had to make sure that the results were not confounded by scrolling. Second, we used comprehension question types that are widely used in recent process model research. The comprehension challenge with BPMN models is to correctly understand the control flow between different activities, i.e. their temporal and logical constraints. *Control flow* is the key mechanism to describe processes and differentiates process models from other forms of modeling, such as object structures or data relationships. The types of questions we use are based on binary relationships such as “Activity a can never be executed before b”. These relationships play an important role for reading, modifying, and validating the model. Typical relationships between a pair of activities are sequential execution order, exclusiveness, concurrency, and repetition. These relationships can be formalized and verified using *behavioral profiles* [33]. Most of the studies to date have used these questions as comprehension measurement instruments, notably because they allow for an objective measurement of control flow comprehension, e.g., [2,3,20].

In our study, we focused on behavioral constraints that can be derived from the structure of the process model. We intentionally ignored domain content, which is typically presented as text annotations of different model elements. The advantage is that there is an objective basis for judging process model comprehension, while confounding effects of domain knowledge can be eliminated [2]. Also, a recent study shows

that control flow comprehension is hindered by the presence of domain information [3], which would have masked some of the effects and results that we are interested in this work.

### 3.4. Measurement

In the experiment, we recorded several variables. Each data point is an observation about a participant answering a comprehension question on one process model. We first use variables to capture the *Model Complexity* and the *Personal Knowledge* with BPMN:

- *Elements*: This variable measures the complexity of the model in terms of the number of gateways in the model.
- *Familiarity*: This variable is an average of the self-evaluation regarding: familiarity with BPMN, familiarity with reading BPMN models, and the number of models read in the last 12 months (the latter divided to 40 to normalize it to an interval of 1 to 5).

Then, we recorded *Cognition Intensity* variables that use data from the eye-tracker:

- *Total Fixations*: This variable captures how many times the participant's attention was focused while solving the comprehension task. It is measured as the count of eye pauses on model elements.
- *Total Duration of Fixations*: This variable captures the total time spent fixating model elements by adding the duration of each fixation. This variable is measured in seconds.

To measure how efficient the subjects investigate a model, we need to associate the fixations' count and duration with the Relevant Region. This focus is in line with the design reported in [34]. *Visual Cognition Efficiency* is measured in line with our previous work [25]:

- *Scan Path Precision*: This variable captures the percentage of fixations on relevant elements in relation to all fixations. In this way, we measure the degree to which participants got distracted by other elements.
- *Scan Path Recall*: This variable captures the percentage of relevant elements that were fixated divided by the number of elements that are relevant for the particular task. In this way, we measure that participants did not miss inspecting relevant elements.

To develop measurements for the dependent variable of *conceptual model comprehension performance*, we automatically recorded the number of correct answers for each of the comprehension tasks. This provided a measure for *comprehension accuracy*. Each of the comprehension questions had an objectively correct answer, which could be answered based on the behavioral semantics of the process model [33]. Accordingly, we define the following dependent variables:

- *Correctness*: This variable captures model comprehension accuracy. It is set to 1 if the participant solves a task correctly and 0 if the provided answer is wrong.
- *Duration*: This variable captures the time for completing a specific comprehension task. This variable is measured in seconds.
- *Efficiency*: This variable divides performance by duration.

## 4. Results

This section presents the results of our experiment, following the recommendations of [35]. First, we provide an overview of the data by summarizing descriptive statistics and screen it for correlations. Second, for testing H1 and H2, we evaluate regression models linking Expertise and Model independent variables to Visual Cognition. Then, Hypothesis H3 is tested by comparing the explanatory power of regression models including the state-of-the-art literature model factors (i.e. Elements and Familiarity) to our alternative regression models based on Visual

**Table 2**  
Summary statistics.

	N	Mean	SD	Minimum	Maximum	Normal Distr.
Familiarity	706	2.633	0.9445	0.667	4.583	No, <0.0001
Elements	706	20.458	8.9024	6.000	33.000	No, <0.0001
Scan Path Precision	706	0.290	0.1448	0.000	1.000	No, <0.0001
Scan Path Recall	706	0.665	0.2365	0.000	1.000	No, <0.0001
Total Duration of Fixations	706	12.643	10.6424	0.000	56.910	No, <0.0001
Total Fixations	706	60.429	47.1881	0.000	271.000	No, <0.0001
Correctness	706	0.865	0.3415	0.000	1.000	No, <0.0001
Efficiency	706	0.0497	0.04736	0.000	0.357	No, <0.0001
Task Duration	706	27.550	18.7011	2.800	115.840	No, <0.0001

Cognition factors (i.e. Total Fixations, Total Duration of Fixations, SPP, and SPR). Finally, we test if Visual Cognition mediates between Personal Knowledge (H4) and Model Complexity (H5) on the one hand, and Comprehension Performance on the other.

#### 4.1. Descriptive statistics and correlation analysis

The data was recorded and processed using Mirametrix EyeMetrix and then imported into SPSS for statistical analysis. First, the observations were filtered such that only the ones that show <10% missing coordinates were kept. Missing coordinates may show up because the subject looked outside the screen or the eye-tracker lost track of the subject's eyes (e.g. if the participant moves her head swiftly). Furthermore, we eliminated all observations connected to an outlier in terms of duration, because the participant took four times as long as the second slowest. Altogether, filtering reduced the number of observations from 864 (72 subjects with 12 questions each) to 706. Note that each observation represents one comprehension question answered by one participant based on a single model. Table 2 summarizes the mean, standard deviation, and range of all variables.

An interesting fact that needs to be pointed out is the high percentage of correct answers (87%), which will impact our further analysis. This high share confirms our selection that the participants are truly experts. The high share of the Relevant Region elements that were actually fixated (the average Scan-path Recall is 66%) is in line with previous research [25].

Table 3 shows the correlation matrix. Note that the dependent variables are moderately correlated with most of the independent and cognitive variables. An exception is Correctness, which does not correlate to other metrics due the high share of correct answers. Duration and, subsequently, Efficiency are correlated with the treatment variables. The former is strongly correlated with Total Duration of Fixations and Total Fixations. This is no surprise considering that a longer examination of the model translates into more and longer fixations. This observation

is likely to hold true for experts since they will be able to focus mostly on the Relevant Region elements. On the other hand, novices would likely investigate the entire model such that Scan Path Precision would drop. All these correlations are in line with previous research on model understanding [25] and with our expectations.

#### 4.2. Hypotheses 1 and 2: Explaining visual cognition

In this section, we test to which degree the visual cognition facets can be explained by the independent variables related to personal knowledge and model complexity. Hypothesis H1 formulates the proposition that Visual Cognition Efficiency can be explained by personal knowledge and model complexity (Table 4). The explanatory power in terms of Adjusted  $R^2$  is, however, quite low. Hypothesis H2 posits that Visual Cognition Intensity can be explained by personal knowledge and model complexity (Table 4). The multivariate regression models of Table 4 confirm this connection, albeit it with a low Adjusted  $R^2$ . Note that in the model for Scan Path Recall, the coefficient for Familiarity points into an unexpected direction.

#### 4.3. Hypothesis 3: Explaining performance

In this section, we test to which degree the independent variables contribute to the explanation of variance of performance. The state-of-the-art model relies on Model Complexity in terms of Elements and Reader Expertise in terms of Familiarity. Our model proposes a direct impact of Visual Cognition (i.e. Total Fixations, Total Duration of Fixations, Scan Path Precision and Recall) on Performance.

We use a forward-selection regression analysis with a stepwise introduction of variables based on log-likelihood. For the binary dependent variable Correctness, we use logistic regression [36]. For the continuous dependent variables Duration and Efficiency, we use linear regression. Each regression model is assessed based on the set of variables and its explanatory power as represented by Nagelkerke's  $R^2$  for

**Table 3**  
Variable correlation matrix.

	Familiarity	Elements	Scan Path Precision	Scan Path Recall	Total Duration of Fixations	Total Fixations	Correctness	Efficiency
Elements	0.013 0.7250							
Scan Path Precision	0.090 0.0170	−0.134 0.0004						
Scan Path Recall	−0.208 < 0.0001	−0.099 0.0086	0.141 0.0002					
Total Duration of Fixations	−0.320 < 0.0001	0.179 < 0.0001	−0.278 < 0.0001	0.486 < 0.0001				
Total Fixations	−0.284 < 0.0001	0.198 < 0.0001	−0.292 < 0.0001	0.497 < 0.0001	0.980 < 0.0001			
Correctness	0.162 < 0.0001	0.098 0.0090	0.062 0.1009	0.083 0.0265	−0.051 0.1768	−0.041 0.2741		
Efficiency	0.296 < 0.0001	−0.148 < 0.0001	0.352 < 0.0001	−0.350 < 0.0001	−0.519 < 0.0001	−0.538 < 0.0001	0.414 < 0.0001	
Task Duration	−0.215 < 0.0001	0.229 < 0.0001	−0.325 < 0.0001	0.394 < 0.0001	0.883 < 0.0001	0.927 < 0.0001	−0.019 0.6209	−0.570 < 0.0001

**Table 4**

Multivariate regression models for Visual Cognition Efficiency in terms of a) Total Fixations and b) Total Duration of Fixations and Visual Cognition Intensity in terms of Scan Path Precision and Scan Path Recall.

Dependent	Independent	Beta	S.E.	t-test	Sig.
Total Fixations	Constant	76.2617			
	Elements	1.0699	1.7664	8.111	<0.0001
	Familiarity	−14.3266	0.1874	5.709	<0.0001
	F = 48.5828	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.1189	
Total Duration of Fixations	Constant	17.7203			
	Elements	0.2194	0.0419	5.235	<0.0001
	Familiarity	−3.6332	0.3950	9.197	<0.0001
	F = 55.3630	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.1336	
Scan Path Precision	Constant	0.2985			
	Elements	−0.0022	0.0006	3.633	<0.0003
	Familiarity	0.0140	0.0060	2.460	<0.0014
	F = 9.5107	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.0236	
Scan Path Recall	Constant	0.8533			
	Elements	−0.0026	0.0010	2.616	<0.0091
	Familiarity	−0.0518	0.0092	5.639	<0.0001
	F = 19.5188	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.0499	

logistic regression or Adjusted R<sup>2</sup> for linear regression. For comparing the explanatory power of the two non-nested models we rely on Steiger's Z-test with 1.96 as the threshold for significance at 95% CI [37].

Table 5 shows that the power of Familiarity plus Elements and cognition-based measures in explaining Correctness is low. This is linked to the little variation observed for the Boolean variable Correctness. As shown in Table 2, there were 86.5% correct answers. Thus, the explanatory power of the logistic regression line is poor. We calculated Steiger's Z value at 1.67 ( $p < 0.001$ ), making the Nagelkerke's R<sup>2</sup> difference of 0.0325 between the two models close to the threshold, on the insignificant side.

For the dependent variable Duration, Table 6 shows a high explanatory power of cognitive variables, which outperform Familiarity plus Elements drastically. All signs point in the expected direction. Scan Path Recall, Scan Path Precision, and the Total Duration of Fixations are negatively connected with Task Duration, while the number of Fixations is positively connected. For example, this means that longer fixations appear to result in lower Task Duration. The Adjusted R<sup>2</sup> of 0.8817 signals a high explanatory power, with a difference of 0.7842 between the two models. We determined Steiger's Z value at 28.8 ( $p < 0.001$ ).

Table 7 shows the regression models for Efficiency dependent variable, for both sets of independent variables. The explanatory power of the cognitive variables is in the mid-range. Higher Total Fixations go with lower Efficiency, while higher Scan Path Precision and Recall are associated with higher Efficiency. We calculated Steiger's Z value, which is 7.37 ( $p < 0.001$ ). Thus, the difference between the Adjusted R<sup>2</sup> of 0.2508 of the two models is significant.

#### 4.4. Hypotheses 4 and 5: Testing for mediation effect

In this section, we investigate hypotheses H4 and H5. We conduct mediation analysis that evaluates if there is a direct relationship

**Table 5**

Logistic regression models considering a) Model and Familiarity variables and b) Visual Cognition variables.

Dependent	Independent	Beta	S.E.	Wald-Sig.	p
Correctness	Constant	−0.034027	0.38529	0.007800	0.9296
	Familiarity	0.52082	0.12344	17.8019	<0.0001
	Elements	0.031210	0.012199	6.5456	0.0105
	Chi <sup>2</sup> = 25.340	Sig. < 0.0001	Nagelkerke R <sup>2</sup>	= 0.06456	
Correctness	Constant	1.15019	0.30417	14.2995	0.0002
	Total Duration of Fixations	−0.031422	0.011006	8.1514	0.0043
	Scan Path Recall	1.74891	0.53631	10.6342	0.0011
	Chi <sup>2</sup> = 10.3906	Sig. = 0.002	Nagelkerke R <sup>2</sup>	= 0.03198	

**Table 6**

Multivariate regression models for Duration considering a) Model and Familiarity variables and b) Visual Cognition variables.

Dependent	Independent	Beta	S.E.	t-test	Sig.
Duration	Constant	28.9640			
	Elements	0.4871	0.07517	6.481	<0.0001
	Familiarity	−4.3221	0.7085	−6.100	<0.0001
	F = 39.09055	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.09752	
Duration	Constant	9.0507			
	Scan Path Precision	−3.7742	1.8654	−2.023	0.0434
	Scan Path Recall	−6.1879	1.2576	−4.921	<0.0001
	Total Duration of Fixations	−1.0954	0.1134	−9.657	<0.0001
	Total Fixations	0.6215	0.02604	23.867	<0.0001
	F = 131,314.89	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.8817	

between the independent I and the dependent Y variable, or whether there is a third variable (the mediator M) such that  $I \rightarrow M \rightarrow Y$ . We look at Model Complexity (namely Elements) and Personal Knowledge (namely Familiarity) as independent variables. The dependent variables Y are the three comprehension performance measures: Correctness, Task Duration, and Efficiency. The Visual Cognition variables are treated as intermediate variables, which might mediate between the Elements and Familiarity variables, respectively, and the dependent variables.

We evaluate the indirect mediation effect according to [38]. The mediation analysis is a two-step approach in which we calculate:

- a multivariate regression (MR) involving both the independent I and the mediating variable M (Fig. 5A). This results in a regression model  $Y_{MR} = c_{MR} + \beta_{MR} * I + \gamma_{MR} * M$ , where  $c_{MR}$  is the intercept of the model, and  $\beta_{MR}$  is the partial-order coefficient for the independent variable I. This model (Fig. 5A) represents the relationship between the dependent and the independent variables after controlling for the effect of the mediating variable.
- a simple regression (SR) involving only the indirect variable I (Fig. 5B). This results in a regression model  $Y_{SR} = c_{SR} + \beta_{SR} * I$ , where  $c_{SR}$  is the intercept of the model, and  $\beta_{SR}$  is the first-order coefficient for the independent variable I. This model represents the direct relationship between the dependent and the independent variables.

The mediation effect is calculated as a difference between the zero-order coefficient from SR and the partial-order coefficient from step MR. Thus  $\beta_{indirect} = \beta_{SR} - \beta_{MR}$ .

For testing the intensity of the mediation effect, we also calculate a simple regression (SM) involving only the mediating variable. This results in a regression model  $M_{SM} = c_{SM} + \beta_{SM} * I$ , where  $c_{SM}$  is the intercept of the model, and  $\beta_{SM}$  is the first-order coefficient for the independent variable I. This model represents the direct relationship between the mediating and the independent variables.

The statistical significance of the indirect effect is determined by the Sobel test. It uses the magnitude of the indirect effect  $\beta_{indirect}$  and compares it to its estimated standard error SE such that  $t = \beta_{indirect} / SE$ ,

**Table 7**

Multivariate regression models for Efficiency considering a) Model and Familiarity variables and b) Visual Cognition variables.

Dependent	Independent	Beta	S.E.	t-test	Sig.
Efficiency	Constant	0.02689			
	Elements	−0.0008073	0.0001892	−4.266	<0.0001
	Familiarity	0.01494	0.001784	8.378	<0.0001
	F = 43.7268	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.1081	
Efficiency	Constant	0.07302			
	Scan Path Precision	0.09170	0.01099	8.347	<0.0001
	Scan Path Recall	0.04335	0.007413	−5.848	<0.0001
	Total Fixations	−0.0003496	0.00003846	−9.091	<0.0001
	F = 132.58	Sig. < 0.001	Adjusted R <sup>2</sup>	= 0.3589	

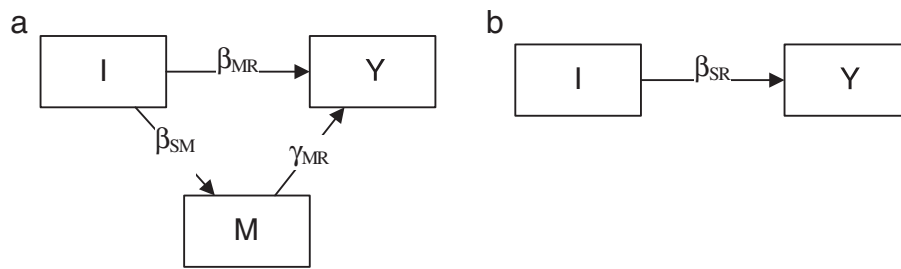


Fig. 5. Mediation analysis approach.

where  $SE = \sqrt{(\beta_{SM}^2 * Var_{\beta_{SM}} + \gamma_{MR}^2 * Var_{\gamma_{MR}})}$  while  $Var_{\beta_{SM}}$  is the variance of  $\beta_{SM}$  and  $Var_{\gamma_{MR}}$  is the variance of  $\gamma_{MR}$ . A significant mediation effect is present if  $\beta_{indirect} > 1.96$ , for  $p = 0.05$ . The Sobel test is applicable to our analysis given that we have a high number of observations, and that the number of observations is the same for each equation.

The results are shown in Table 8. It can be seen that the relationship between all three independent variables and all four dependent variables are significantly mediated both by the Visual Cognition Intensity

and Visual Cognition Efficiency variables. This is indicated by the Sobel test, having a value  $> 1.96$ .

However, not all mediation relationships are equally strong. The mediation between all dependent variables and independent Correctness is statistically relevant as indicated by Sobel test, but the low  $\beta_{indirect}$  (it is at most 10% compared to the  $\beta_{SR}$ ) added to the very low Nagelkerke's  $R^2$  of the logistic regression suggests that there can be other variables that influence the ability of the readers to provide

**Table 8**  
Mediation test (considering all possible combinations of variables).

Dependent variable (Y)	Independent variable (I)	Mediating variable (M)	MR	SR	SM	$B_{indirect}$	Sobel test (p value)
Correctness	Elements	Total Fixations	$Y = 0.806 + 0.0042 I + (-0.0005) M$	$Y = 0.7884 + 0.0037 I$	$M = 38.954 + 1049 I$	-0.0005	5202.4 (<0.01)
		Total Duration of Fixations	$Y = 0.807 + 0.0042 I + (-0.0023) M$	$Y = 0.8835 + 0.0037 I$	$M = 18.561 + 0.149 I$	-0.0005	7968.7 (<0.01)
		Scan Path Precision	$Y = 0.728 + 0.0041 I + 0.18 M$	$Y = 0.8835 + 0.0037 I$	$M = 0.335 + (-0.002) I$	-0.0004	8304.9 (<0.01)
		Scan Path Recall	$Y = 0.691 + 0.0041 I + 0.136 M$	$Y = 0.8835 + 0.0037 I$	$M = 0.718 + (-0.002) I$	-0.0004	10,172 (<0.01)
Correctness	Familiarity	Total Fixations	$Y = 0.707 + 0.05916 I + 0.00004 M$	$Y = 0.711 + 0.05862 I$	$M = 97.797 + (-14.19) I$	0.00054	52.82 (<0.01)
		Total Duration of Fixations	$Y = 0.710 + 0.05875 I + 0.000003 M$	$Y = 0.711 + 0.05862 I$	$M = 22.136 + (-3605) I$	-0.00013	8.32 (<0.01)
		Scan Path Precision	$Y = 0.683 + 0.057 I + 0.112 M$	$Y = 0.711 + 0.059 I$	$M = 0.254 + 0.013 I$	-0.002	897.5 (<0.01)
		Scan Path Recall	$Y = 0.569 + 0.068 I + 0.177 M$	$Y = 0.711 + 0.059 I$	$M = 0.802 + (-0.052) I$	-0.009	883.1 (<0.01)
Task Duration	Elements	Total Fixations	$Y = 3.539 + 0.099 I + 0.364 M$	$Y = 5.344 + 0.368 I$	$M = 38.954 + 1049 I$	0.269	75.58 (<0.01)
		Total Duration of Fixations	$Y = 5.084 + 0.154 I + 1.527 M$	$Y = 5.344 + 0.368 I$	$M = 18.561 + 0.149 I$	0.214	100.5 (<0.01)
		Scan Path Precision	$Y = 30.691 + 0.397 I + (-38.751) M$	$Y = 5.344 + 0.368 I$	$M = 0.335 + (-0.002) I$	-0.029	124.3 (<0.01)
		Scan Path Recall	$Y = -6.171 + 0.568 I + 33.245 M$	$Y = 5.344 + 0.368 I$	$M = 0.718 + (-0.002) I$	-0.200	72.04 (<0.01)
Task Duration	Familiarity	Total Fixations	$Y = 2.254 + 1.038 I + 0.373 M$	$Y = 38.769 + (-4.261) I$	$M = 97.797 + (-14.19) I$	-5.299	11.55 (<0.01)
		Total Duration of Fixations	$Y = 3.488 + 1.486 I + 1.594 M$	$Y = 38.769 + (-4.261) I$	$M = 22.136 + (-3605) I$	-5.747	13.56 (<0.01)
		Scan Path Precision	$Y = 48.89 + (-3.713) I + (-39.8) M$	$Y = 38.769 + (-4.261) I$	$M = 0.254 + 0.013 I$	-0.548	12.01 (<0.01)
		Scan Path Recall	$Y = 15.64 + (-2.757) I + 28.838 M$	$Y = 38.769 + (-4.261) I$	$M = 0.802 + (-0.052) I$	-1.504	17.01 (<0.01)
Efficiency	Elements	Total Fixations	$Y = 0.08 + (-0.0002) I + (-0.0005) M$	$Y = 0.066 + 0.00079 I$	$M = 38.954 + 1049 I$	0.001	49,968 (<0.01)
		Total Duration of Fixations	$Y = 0.08 + (-0.0003) I + (-0.002) M$	$Y = 0.066 + 0.00079 I$	$M = 18.561 + 0.149 I$	0.0011	73,290 (<0.01)
		Scan Path Precision	$Y = 0.03 + (-0.0005) I + 0.11 M$	$Y = 0.066 + 0.00079 I$	$M = 0.335 + (-0.002) I$	0.0013	45,454 (<0.01)
		Scan Path Recall	$Y = 0.12 + (-0.001) I + (-0.073) M$	$Y = 0.066 + 0.00079 I$	$M = 0.718 + (-0.002) I$	0.0018	32,731 (<0.01)
Efficiency	Familiarity	Total Fixations	$Y = 0.059 + 0.008 I + (-0.0005) M$	$Y = 0.011 + 0.015 I$	$M = 97.797 + (-14.19) I$	0.007	8080 (<0.01)
		Total Duration of Fixations	$Y = 0.057 + 0.007 I + (-0.002) M$	$Y = 0.011 + 0.015 I$	$M = 22.136 + (-3605) I$	0.008	9286 (<0.01)
		Scan Path Precision	$Y = -0.017 + 0.014 I + 0.107 M$	$Y = 0.011 + 0.015 I$	$M = 0.254 + 0.013 I$	0.001	4535 (<0.01)
		Scan Path Recall	$Y = 0.059 + 0.012 I + (-0.060) M$	$Y = 0.011 + 0.015 I$	$M = 0.802 + (-0.052) I$	0.003	7356 (<0.01)



correct answers. The mediation between the dependent Task Duration and independent Elements and Familiarity is very strong for Visual Cognition Intensity variables, as indicated by the high  $\beta_{\text{indirect}}$  and Sobel test values. For Visual Cognition Efficiency, the mediation effect is of medium strength for Scan Path Recall and weak for Scan Path Precision. The mediation between the dependent Efficiency and both independent variables is statistically relevant for all mediating variables, as indicated by the very high Sobel test, but  $\beta_{\text{indirect}}$  is marginal.

Overall, Table 8 confirms our expectation that the size of the model and expertise of the model reader are not directly influencing the correctness of answering comprehension questions or how long it takes the reader to make sense of the model. Rather, it influences the cognitive effort to find and focus on the relevant elements for the task at hand, which subsequently determines comprehension performance.

## 5. Discussion

This section discusses the implications of our research findings. Section 5.1 summarizes the results. Section 5.2 discusses implications for research and practice. Section 5.3 clarifies potential threats to the validity of our study.

### 5.1. Summary of results

We set up this study to evaluate the hypotheses of our research model and the difference between the explanatory power compared the state-of-the-art Literature Model and the Visual Cognition Model. We hypothesized that the reason for a positive effect on comprehension performance would stem from an improvement in visual cognition. Process model understanding performance was measured as: correctness and duration of answering comprehension questions, as well as efficiency, which considers both.

Table 9 summarizes the results. Both hypotheses H1 and H2 are supported, with the minor limitation of one coefficient for Scan Path Precision pointing in an unexpected direction. Low explanatory power suggests the presence of supplementary factors. Regarding H3, the regression models for Correctness provide a weak explanation for performance variation. We find that the hypothesis is strongly supported in the two performance dimensions of Duration and Efficiency. We observe that the Visual Cognition regression model for Duration has a very high explanatory power. By contrast, the regression model for Duration based on Model Complexity and Familiarity seems unable to explain the Task Duration. Efficiency, which involves both Correctness and Duration, also points out the superiority of the Visual Cognition model.

Hypotheses H4 and H5 are supported for all performance dimensions. Therefore, we find support for our claim that there is no direct influence between reader Familiarity and model Complexity on the one

side and model reading Performance on the other. The strongest mediation effect suggests that the Duration of the task has little to do with Familiarity or Model Size, but is determined by Visual Cognition. The direct implication of this finding is that reading performance could be increased if Visual Comprehension Intensity and Efficiency are improved. This may not necessarily require an increase the Familiarity/Expertise of the reader or a decrease of the Model Size.

### 5.2. Implications for research and practice

The findings reported in this paper have several implications for research and practice. We will start with a discussion of the three implications for research that we see. First, the results reported in this paper are highly relevant for experimental research on process model understanding. Many of the prior works consider size and complexity metrics to be relevant for model understanding, e.g. [1,13–15,20]. What we find in this paper, however, are results that strongly emphasize the complexity of the particular understanding task. Obviously, a task that refers to a small relevant region of a complex model appears to be easier to solve than a large relevant region in a relatively small model. This finding is only partially acknowledged in prior research [15]. The distinction between the complexity of the overall model on the one hand and the problem-solving task on the other might also lead to a more cautious and more specific usage of the terms understanding and understandability in experimental research on conceptual models. *Understanding* in this sense can be regarded as being anchored in a specific task, while *understandability* apparently refers to the entirety of matters that can be understood from a model. Such a distinction might also help to provide a better justification for the selection of understanding tasks in experiments based on their implied relevant region. In this way, our research contributes to a better internal validity of future experiments in this area.

Second, our research emphasizes the merits of eye-tracking for investigating decision making in conceptual modeling. A systematic literature review shows that eye-tracking research on conceptual modeling is still at its infancy. The few studies rather explore correlations instead of testing theoretical models [39]. The variables of visual cognition that we propose help to overcome the metrics confusion that is reported in this systematic literature review. Most prior studies on process model comprehension relied on performance data alone. The different regression models that we estimated demonstrate the benefits of taking variables of visual cognition into account. Eye-tracking equipment might be specifically helpful to investigate how people create models throughout the process of process modeling [40].

Our study also provides insights into the extent at which visual cognition variables explain different aspects of comprehension performance. The explanatory power appears to be high for Task Duration

**Table 9**  
Results summary.

Hypothesis	Performance dimension	Explanatory power	Support
H1	Total Fixations	0.1189	Supported
H1	Total Duration of Fixations	0.1336	Supported
H2	Scan Path Precision	0.0236	Partially supported
H2	Scan Path Recall	0.0499	Supported
H3	Correctness	Increase in Explanatory power: difference of $R^2$ Visual Cognition model– $R^2$ Literature model –0.033 = (0.032–0.065)	Not supported
H3	Duration	Increase in Explanatory power difference 0.782 = (0.882–0.100)	Supported
H3	Efficiency	Increase in Explanatory power difference 0.249 = (0.359–0.110)	Supported
H4	Correctness	Familiarity mediated by Total Fixations, Total Duration of Fixations, SPP, and SPR (each $p < 0.01$ )	Supported
H4	Duration	Familiarity mediated by Total Fixations, Total Duration of Fixations, SPP, and SPR (each $p < 0.01$ )	Supported
H4	Efficiency	Familiarity mediated by Total Fixations, Total Duration of Fixations, SPP, and SPR (each $p < 0.01$ )	Supported
H5	Correctness	Elements mediated by Total Fixations, Total Duration of Fixations, SPP, and SPR (each $p < 0.01$ )	Supported
H5	Duration	Elements mediated by Total Fixations, Total Duration of Fixations, SPP, and SPR (each $p < 0.01$ )	Supported
H5	Efficiency	Elements mediated by Total Fixations, Total Duration of Fixations, SPP, and SPR (each $p < 0.01$ )	Supported

and efficiency while the contribution to correctness is small. Presumably, the correctness of answers might be stronger influenced by schematic knowledge in the long-term memory of a model reader. Visual cognition, however, appears to be at least closely connected with Task Duration. This suggests that correctness tends to depend on long-term memory knowledge while visual cognition influences how quickly this knowledge can be activated.

Our observations also have implications for practice. First, based on our insights, tool features can be designed to help the model reader in reading and understanding a model. Features for providing visual cues can be designed to highlight relevant regions for a specific comprehension question at hand. Second, modelers can be directed to those parts of their model that are likely to be difficult to understand by others. This would require calculating the biggest relevant region for a model or all relevant regions greater than a certain threshold size. The modeler can then be supported to find refactorings of the model to make it easier to comprehend. For instance, techniques can be used to automatically derive a process model that has a better structure [15]. This can be helpful when a domain expert has to validate a process model. Finally, tracking visual cognition during working with a process model could be extended towards an assessment tool for modeling competence. Currently, certification programs are often based on multiple choice tests. If eye-tracking is used, a certification agency could check if a candidate also looks at the relevant regions in the model, i.e. if the candidate truly understands the model.

### 5.3. Threats to validity

There are potential threats to the validity of our study that need to be mentioned. First, internal validity might be threatened by the imprecision of the eye-tracking system. Several factors influence the precision of the eye-tracker, like eye movement physiology, calibration accuracy, and ambient light. Throughout the course of the experiment, we took special care to calibrate and also re-calibrate to minimize this threat. What is more, we kept only observations containing <10% missing data.

Second, construct validity is concerned with the trustworthiness of the measurements. We use familiarity and complexity measures from prior research. The cognitive measures build on objective data recordings and are, therefore, not subject to concerns of perceptual measures.

Third, conclusion validity in our experiment faces a trade-off between variation in time and variation in correctness. We chose for working with experts, which provided us with a low variance in correct-

ness and good variance in duration. For novices, we would probably have seen a greater variance in correctness and a more uniform albeit longer duration. Our choice is motivated by our interest in visual cognition. The results indeed confirm that cognitive variables have a high explanatory power for duration. As drawback, the large share of correct answers by the expert participants has contributed to partially insignificant results for the Correctness metric.

Fourth, to address external validity concerns, we aimed for a mix of academia and industry participants as well as geographical spread in an effort to ensure a high degree of generality to our findings. Also, we used BPMN models because it is the most widely adopted type of modeling notation in teaching and in industry. Another possible threat to the external validity is the nature of the models. We partially reused models from other studies [3] and partially integrated models derived from industry projects in order to avoid experimenter bias.

## 6. Conclusions

In this paper, we found that visual cognition variables outperform model complexity and personal knowledge, in terms of familiarity, in explaining the variance of Task Duration and efficiency. Furthermore, we find that the statistical effect of model complexity and familiarity is fully mediated by visual cognition variables for all considered performance variables.

Our findings have strong implications for future research on process model comprehension. It emphasizes the relevance of eye-tracking for investigating comprehension performance in the area of conceptual modeling. We found out that improving visual cognition is a better avenue to increase comprehension than increasing expertise or decreasing model complexity. Our insights also suggest the benefits of tool features that help a model reader in understanding a model while directing the modeler to those parts of their model that are likely to be difficult to understand by others. Future research may complement our study, which puts the focus on the consumption side of process models, by incorporating eye-tracking in studies into the production side, i.e. the creation and the process of process modeling.

## Acknowledgement

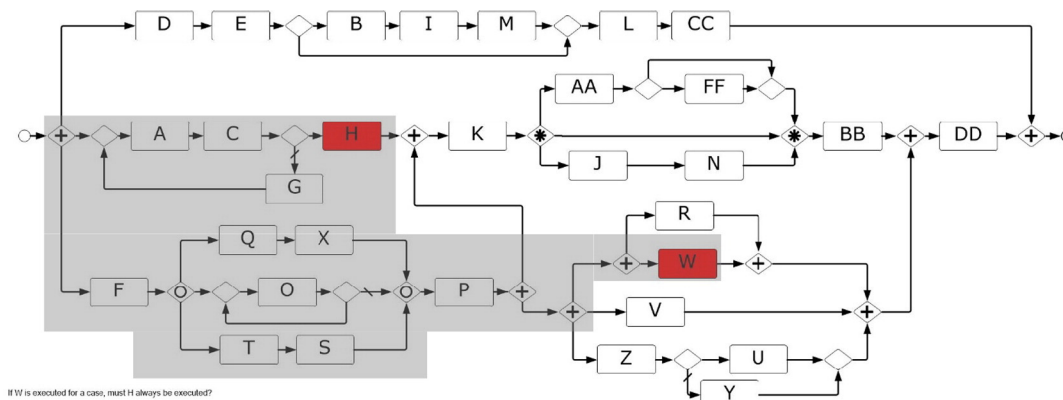
This work was supported by UBB-GTC grant no. 34018/2013. We would also like to thank all the people who participated in this study.

## Appendix A

Note that the Relevant Region elements are shared gray in this exhibit. The subjects did not have any shading in their material.

Question Model 1: If W is executed for a case, must H always be executed?

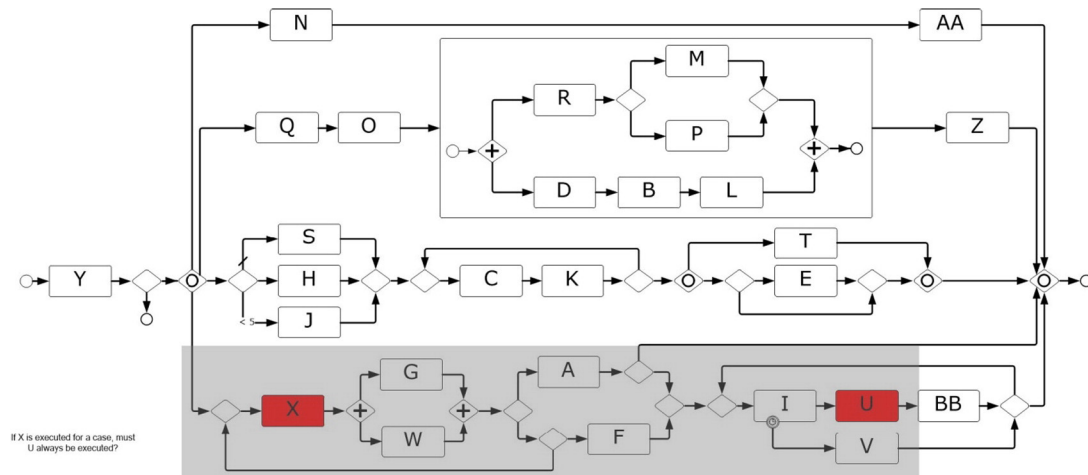
Metrics: Number of model elements: 122; and Number of control-flow elements: 23.



If W is executed for a case, must H always be executed?

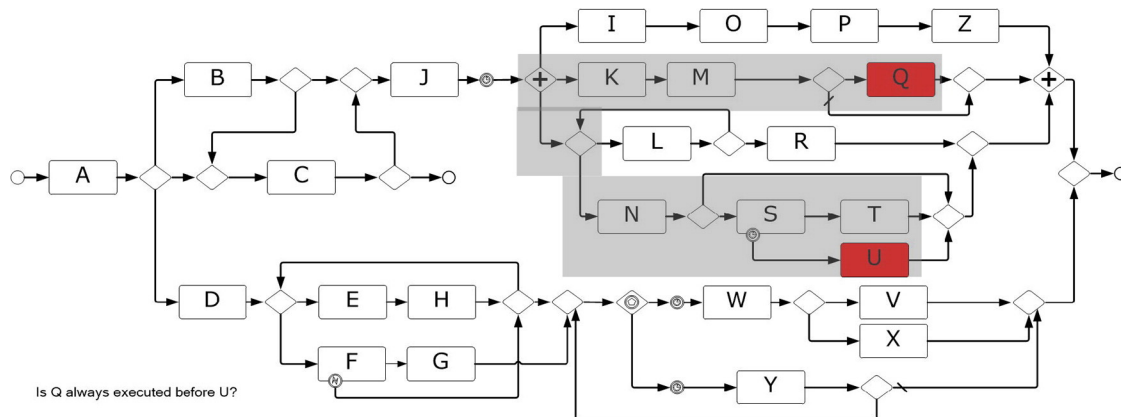
Question Model 2: If X is executed for a case, must U always be executed?

Metrics: Number of model elements: 124; and Number of control-flow elements: 25.



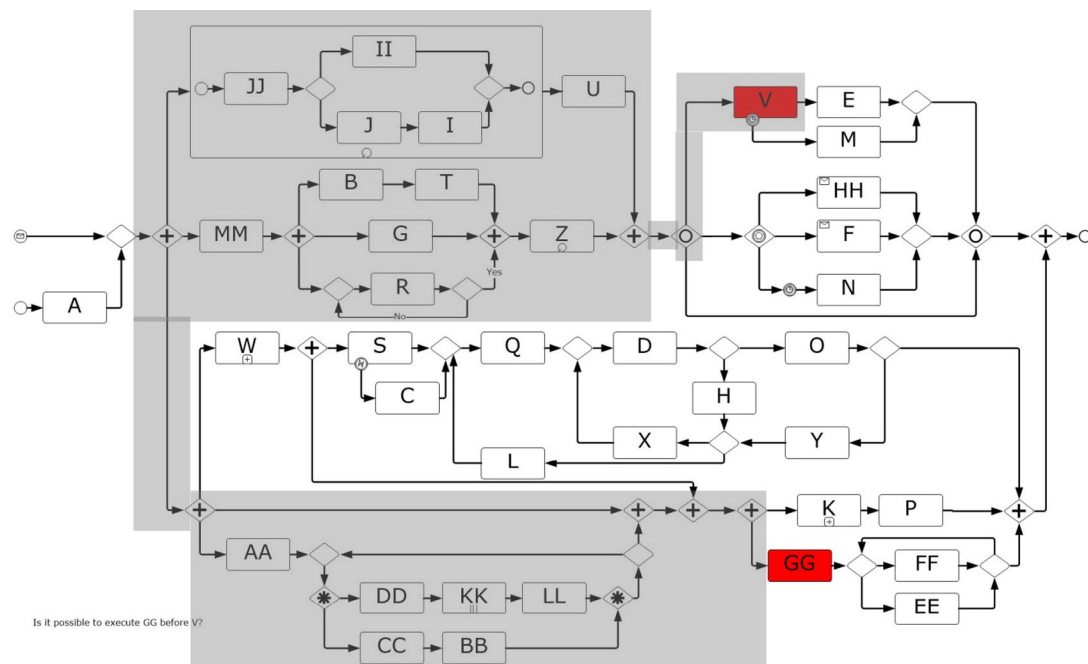
Question Model 3: Is Q always executed before U?

Metrics: Number of model elements: 118; and Number of control-flow elements: 24.



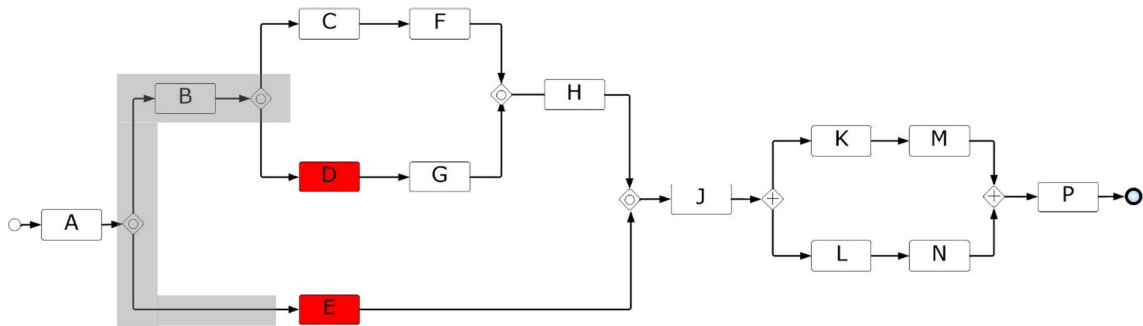
Question Model 4: Is it possible to execute GG before V?

Metrics: Number of model elements: 169; and Number of control-flow elements: 33.



Question Model 5: If D is executed, is it possible to execute E in the same case?

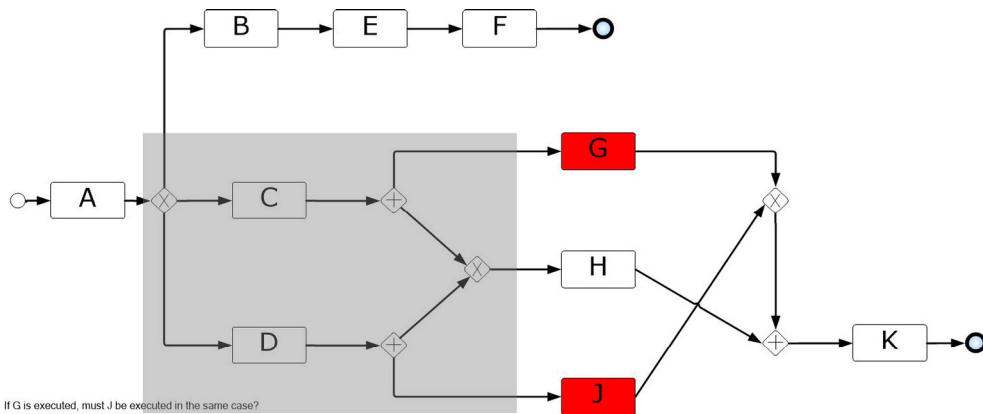
Metrics: Number of model elements: 44; and Number of control-flow elements: 6.



If D is executed, is it possible to execute E in the same case?

Question Model 6: If G is executed, must J be executed in the same case?

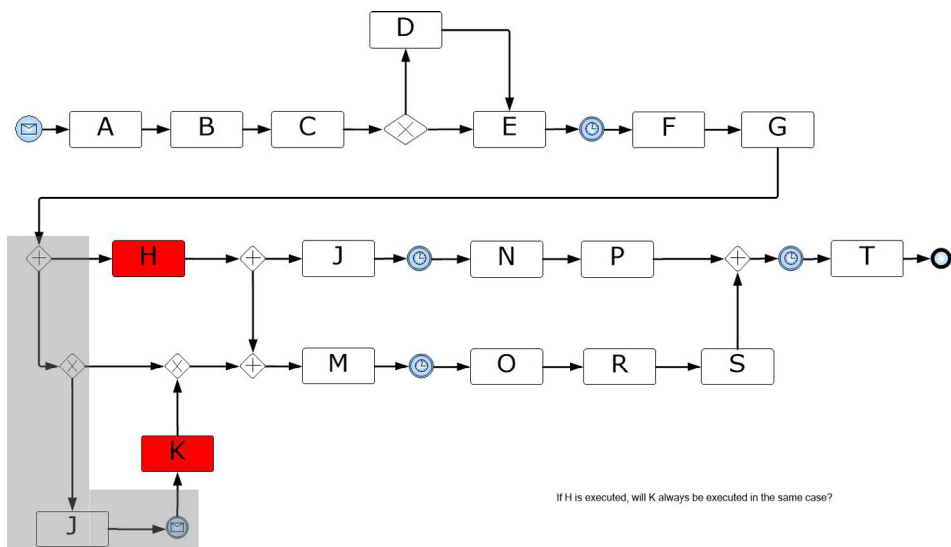
Metrics: Number of model elements: 37; and Number of control-flow elements: 6.



If G is executed, must J be executed in the same case?

Question Model 7: If H is executed, will K always be executed in the same case?

Metrics: Number of model elements: 60; and Number of control-flow elements: 7.

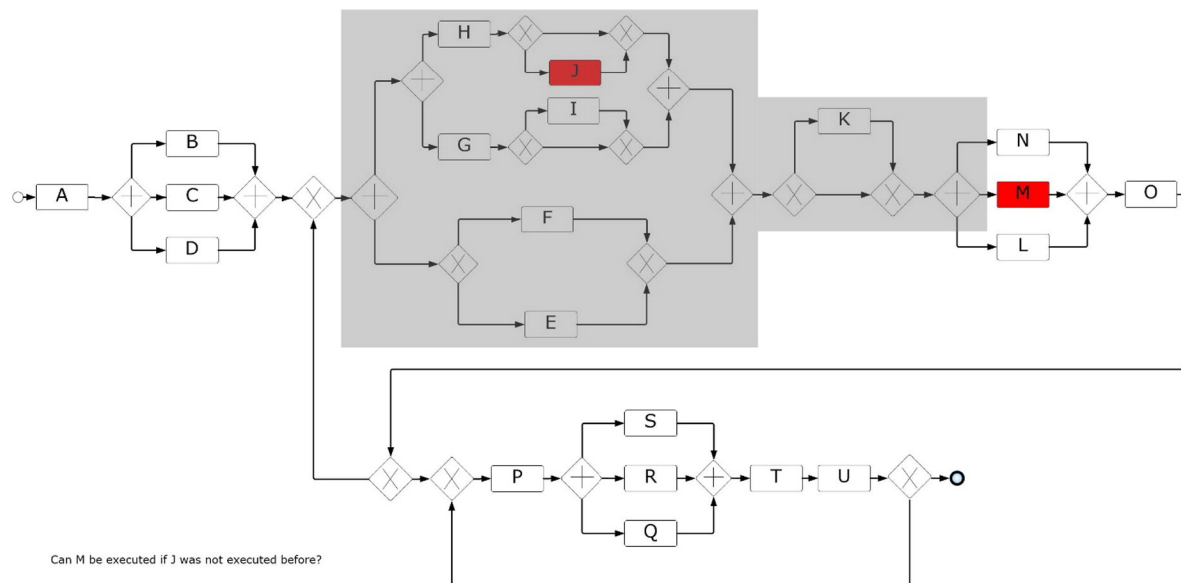


If H is executed, will K always be executed in the same case?



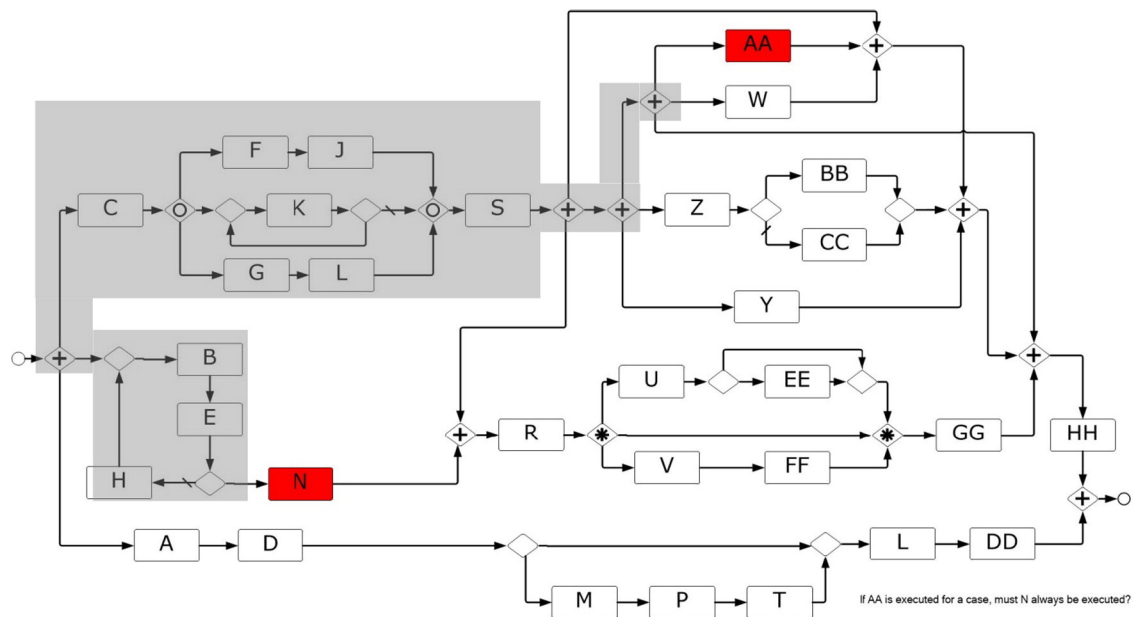
Question Model 8: Can M be executed if J was not executed before?

Metrics: Number of model elements: 101; and Number of control-flow elements: 22.

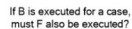


Question Model 9: If AA is executed for a case, must N always be executed?

Metrics: Number of model elements: 123; and Number of control-flow elements: 23.



Metrics: Number of model elements: 123; and Number of control-flow elements: 24.

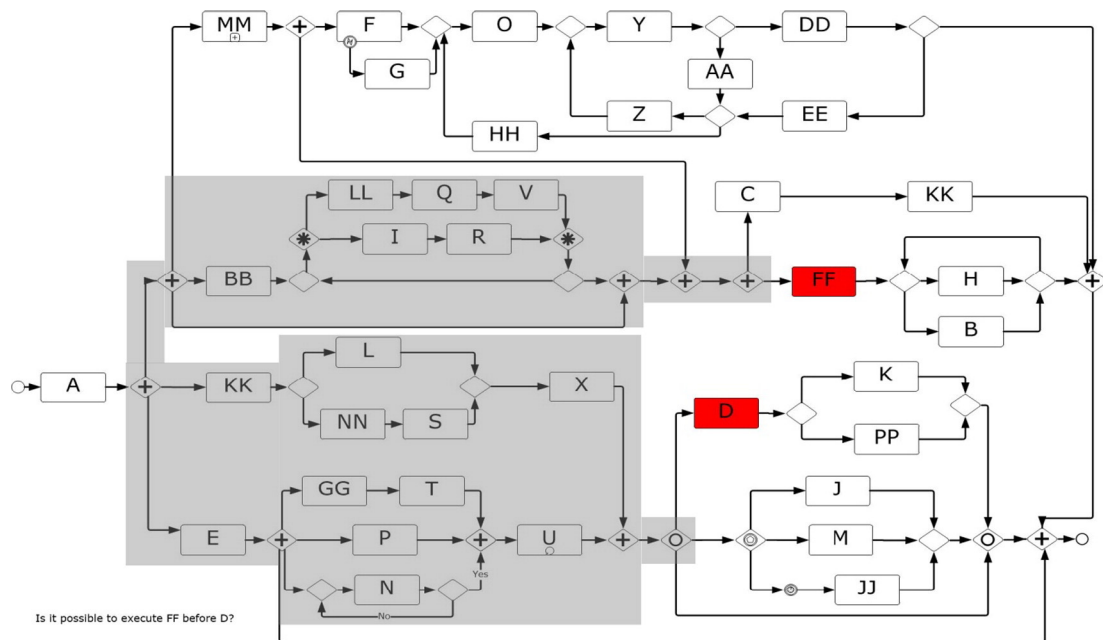


Metrics: Number of model elements: 118; and Number of control-flow elements: 24.



Question Model 12: Is it possible to execute FF before D?

Metrics: Number of model elements: 169; and Number of control-flow elements: 33.



## References

- [1] M. Dumas, M. La Rosa, J. Mendling, H.A. Reijers, Fundamentals of Business Process Management, Springer, 2013 <http://dx.doi.org/10.1007/978-3-642-33143-5>.
- [2] H.A. Reijers, J. Mendling, A study into the factors that influence the understandability of business process models, Syst. Man Cybern. Part A Syst. Hum. IEEE Trans. 41 (3) (2011) 449–462, <http://dx.doi.org/10.1109/tsmca.2010.2087017>.
- [3] J. Mendling, M. Strembeck, J. Recker, Factors of process model comprehension—findings from a series of experiments, Decis. Support. Syst. 53 (1) (2012) 195–206, <http://dx.doi.org/10.1016/j.dss.2011.12.013>.
- [4] J.C. Recker, A. Dreiling, The effects of content presentation format and user characteristics on novice developers' understanding of process models, CAIS 28 (6) (2011) 65–84.
- [5] J. Recker, Continued use of process modeling grammars: the impact of individual difference factors, Eur. J. Inf. Syst. 19 (1) (2010) 76–92, <http://dx.doi.org/10.1057/ejis.2010.5>.
- [6] A. Burton-Jones, P.N. Meso, The effects of decomposition quality and multiple forms of information on novices' understanding of a domain from a conceptual model, JAIS 9 (12) (2008) 748.
- [7] K. Figl, R. Laue, Influence factors for local comprehensibility of process models, Int. J. Hum. Comput. Stud. 82 (2015) 96–110, <http://dx.doi.org/10.1016/j.ijhcs.2015.05.007>.
- [8] H.L. Fromkin, S. Streufert, Laboratory Experimentation, No. 343, Herman C. Krannert Graduate School of Industrial Administration, Purdue University, 1972.
- [9] D.L. Moody, Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions, DKE 55 (3) (2005) 243–276, <http://dx.doi.org/10.1016/j.dataak.2004.12.005>.
- [10] J. Recker, H.A. Reijers, S.G. van de Wouw, Process model comprehension: the effects of cognitive abilities, learning style, and strategy, CAIS 34 (1) (2014) 9.
- [11] A. Gemino, Y. Wand, A framework for empirical evaluation of conceptual modeling techniques, Requir. Eng. 9 (4) (2004) 248–260, <http://dx.doi.org/10.1007/s00766-004-0204-6>.
- [12] J. Mendling, H.A. Reijers, J. Cardoso, What makes process models understandable? Business Process Management, Springer 2007, pp. 48–63, [http://dx.doi.org/10.1007/978-3-540-75183-0\\_4](http://dx.doi.org/10.1007/978-3-540-75183-0_4).
- [13] J. Cardoso, Evaluating workflows and web process complexity, Workflow Handbook 2005 2005, pp. 284–290.
- [14] L. Sanchez Gonzalez, F. Garcia Rubio, F. Ruiz Gonzalez, M. Piattini Velthuis, Measurement in business processes: a systematic review, Bus. Process. Manag. J. 16 (1) (2010) 114–134, <http://dx.doi.org/10.1108/14637151011017976>.
- [15] A. Polyvyanyy, L. Garcá-Bañuelos, M. Dumas, Structuring acyclic process models, Inf. Syst. 37 (6) (2012) 518–538, <http://dx.doi.org/10.1016/j.is.2011.10.005>.
- [16] K. Figl, J. Mendling, M. Strembeck, The influence of notational deficiencies on process model comprehension, J. Assoc. Inf. Syst. 14 (6) (2013) 312.
- [17] J. Recker, M. Rosemann, P.F. Green, M. Indulska, Do ontological deficiencies in modeling grammars matter? MIS Q. 35 (1) (2011) 57–79.
- [18] M. Zur Muehlen, J. Recker, How much language is enough? Theoretical and practical use of the business process modeling notation, CAISE, Springer 2008, pp. 465–479, [http://dx.doi.org/10.1007/978-3-540-69534-9\\_35](http://dx.doi.org/10.1007/978-3-540-69534-9_35).
- [19] P. Bera, Does cognitive overload matter in understanding bpmn models? J. Comput. Inf. Syst. 52 (4) (2012) 59–69.
- [20] H.A. Reijers, T. Freytag, J. Mendling, A. Eckleder, Syntax highlighting in business process models, Decis. Support. Syst. 51 (3) (2011) 339–349, <http://dx.doi.org/10.1016/j.dss.2010.12.013>.
- [21] R. Petrusel, J. Mendling, H.A. Reijers, Task-specific visual cues for improving process model understanding, IST 79 (2016) 63–78, <http://dx.doi.org/10.1016/j.infsof.2016.07.003>.
- [22] J. Mendling, H.A. Reijers, J. Recker, Activity labeling in process modeling: empirical insights and recommendations, Inf. Syst. 35 (4) (2010) 467–482, <http://dx.doi.org/10.1016/j.is.2009.03.009>.
- [23] M.A. Just, P.A. Carpenter, Eye fixations and cognitive processes, Cogn. Psychol. 8 (4) (1976) 441–480, [http://dx.doi.org/10.1016/0010-0285\(76\)90015-3](http://dx.doi.org/10.1016/0010-0285(76)90015-3).
- [24] J.R. Anderson, D. Bothell, S. Douglass, Eye movements do not reflect retrieval processes limits of the eye-mind hypothesis, Psychol. Sci. 15 (4) (2004) 225–231, <http://dx.doi.org/10.1111/j.0956-7976.2004.00656.x>.
- [25] R. Petrusel, J. Mendling, Eye-tracking the factors of process model comprehension tasks, CAISE, Springer 2013, pp. 224–239, [http://dx.doi.org/10.1007/978-3-642-38709-8\\_15](http://dx.doi.org/10.1007/978-3-642-38709-8_15).
- [26] A. Duchowski, Eye Tracking Methodology: Theory and Practice, vol. 373, Springer, 2007.
- [27] K. Holmqvist, M. Nystrom, R. Andersson, R. Dewhurst, H. Jarodzka, J. Van de Weijer, Eye Tracking: A Comprehensive Guide to Methods and Measures, Oxford University Press, 2011.
- [28] R. Baeza-Yates, B. Ribeiro-Neto, et al., Modern Information Retrieval, vol. 463, ACM press, New York, 1999.
- [29] J. Sweller, Cognitive load during problem solving: effects on learning, Cogn. Sci. 12 (2) (1988) 257–285, [http://dx.doi.org/10.1207/s15516709cog1202\\_4](http://dx.doi.org/10.1207/s15516709cog1202_4).
- [30] S.J. Luck, E.K. Vogel, The capacity of visual working memory for features and conjunctions, Nature 390 (6657) (1997) 279–281.
- [31] G.A. Alvarez, Representing multiple objects as an ensemble enhances visual cognition, Trends Cogn. Sci. 15 (3) (2011) 122–131, <http://dx.doi.org/10.1016/j.tics.2011.01.003>.
- [32] H.Y. Eng, D. Chen, Y. Jiang, Visual working memory for simple and complex visual stimuli, Psychon. Bull. Rev. 12 (6) (2005) 1127–1133, <http://dx.doi.org/10.3758/bf03206454>.
- [33] M. Weidlich, J. Mendling, M. Weske, Efficient consistency measurement based on behavioral profiles of process models, IEEE Trans. Softw. Eng. 37 (3) (2011) 410–429, <http://dx.doi.org/10.1109/tse.2010.96>.
- [34] C. Conati, C. Merten, Eye-tracking for user modeling in exploratory learning environments: an empirical evaluation, Knowl.-Based Syst. 20 (6) (2007) 557–574, <http://dx.doi.org/10.1016/j.knosys.2007.04.010>.
- [35] A. Field, Discovering Statistics Using IBM SPSS Statistics, Sage, 2013.
- [36] D.W. Hosmer Jr., S. Lemeshow, Applied Logistic Regression, John Wiley & Sons, 2004, <http://dx.doi.org/10.1002/0471722146>.
- [37] J.H. Steiger, Tests for comparing elements of a correlation matrix, Psychol. Bull. 87 (2) (1980) 245, <http://dx.doi.org/10.1037/0033-2909.87.2.245>.
- [38] C.M. Judd, D.A. Kenny, Process analysis estimating mediation in treatment evaluations, Eval. Rev. 5 (5) (1981) 602–619, <http://dx.doi.org/10.1177/0193841x8100500502>.
- [39] Z. Sharafi, Z. Soh, Y.-G. Gueheneuc, A systematic literature review on the usage of eye-tracking in software engineering, IST 67 (2015) 79–107, <http://dx.doi.org/10.1016/j.infsof.2015.06.008>.
- [40] J. Pinggera, M. Furtner, M. Martini, P. Sachse, K. Reiter, S. Zugal, B. Weber, Investigating the process of process modeling with eye movement analysis, Business Process Management Workshops, Springer 2013, pp. 438–450, [http://dx.doi.org/10.1007/978-3-642-36285-9\\_46](http://dx.doi.org/10.1007/978-3-642-36285-9_46).



**Razvan Petrusel** is an Associate Professor with the Department of Business Information Systems of the Faculty of Economics and Business Administration in Babeş-Bolyai University of Cluj-Napoca, Romania. He received his Ph.D. in Cybernetics and Statistics in 2008 from Babeş-Bolyai University. His research interests, as well as teaching, include Business Process Management, Conceptual Modeling, Decision Modeling and Mining, and Process Mining. He published and presented over 40 research papers in journals (e.g. Information and Software Technology) and at conferences (e.g. CalSE, BIS).



**Jan Mendling** is a Full Professor with the Institute for Information Business at Wirtshschaftsuniversität Wien (WU Vienna), Austria. His research interests include various topics in the area of business process management and information systems. He has published >250 research papers and articles, among others in ACM Transactions on Software Engineering and Methodology, IEEE Transaction on Software Engineering, Information Systems, Data & Knowledge Engineering, and Decision Support Systems. He is member of the editorial board of seven international journals, member of the board of the Austrian Society for Process Management (, <http://prozesse.at>), one of the founders of the Berlin BPM Community of Practice (, <http://www.bpmb.de>), organizer of several academic events on process management, and member of

the IEEE Task Force on Process Mining. His Ph.D. thesis has won the Heinz-Zemanek-Award of the Austrian Computer Society and the German Targion-Award for dissertations in the area of strategic information management.



**Hajo Reijers** is a full professor in Business Informatics at the Department of Computer Science of the Vrije Universiteit Amsterdam (VU), where he is head of the Business Informatics group. Furthermore, he is a part-time full professor in the AIS group of the Department of Mathematics and Computer Science of Eindhoven University of Technology (TU/e). His research and teaching focus on business process management, workflow technology, business process improvement, and conceptual modeling. On these and related topics, he published over 150 research papers, book chapters and professional publications. He is also the managing director of the European BPM Round Table, an initiative to connect researchers and practitioners in the area of Business Process Management («, <http://www.bpmroundtable.eu/>). He is

closely cooperating with companies from the services and healthcare domains, as well as with various international scholars.