

Cognitive Complexity in Business Process Modeling

Kathrin Figl¹ and Ralf Laue²

¹ Vienna University of Economics and Business Administration, Austria
kathrin.figl@wu.ac.at

² Computer Science Faculty, University of Leipzig, Germany
laue@ebus.informatik.uni-leipzig.de

Abstract. Although (business) process models are frequently used to promote human understanding of processes, practice shows that understanding complex models soon reach cognitive limits. The aim of this paper is to investigate the cognitive difficulty of understanding different relations between model elements. To allow for empirical assessment of this research question we systematically constructed model sets and comprehension questions. The results of an empirical study with 199 students tend to suggest that comprehension questions on order and concurrency are easier to answer than on repetition and exclusiveness. Additionally, results lend support to the hypothesis that interactivity of model elements influences cognitive difficulty. While our findings shed light on human comprehension of process models, they also contribute to the question on how to assure understandability of models in practice.

Key words: Business Process Models, Understandability, Cognitive Complexity

1 Introduction

Business process models (BPM) serve as a basis for communication between domain experts, business process analysts and software developers. To fulfill this purpose, such models have to be easy to understand and easy to maintain. Comprehension of process models is relevant for all tasks in which users interact with models, as for example in business process redesign or implementation of process-aware systems.

Many researchers have recently turned to investigate comprehensibility of process models and investigated various influence factors as modularity [1], domain knowledge [2] and notational aspects [3]. In addition, various complexity metrics have been proposed for BPM in the past years (see [4, 5] for the discussion of relevant concepts and [6] for a comprehensive survey on related work). It has been shown that some of these metrics are significantly correlated with the number of control-flow errors in a BPM [5] and with the understandability of a BPM, measured in terms of correctly answered questions about the model [7, 8]. [2] and [7] discuss how global measures (like the number of split nodes in a BPM) affect the understandability of a BPM.

However, the scope of existing studies is limited, because the metrics used in these studies assign single (global) values to a BPM to describe its complexity. Ananda et al. [9] state: “Although studying the overall comprehensibility of a model is important, from a language evolution perspective it is even more relevant to discover *which* elements of a notation work well and which do not.” With this paper, we want to give some first answers on the question *which* relations between model elements in a BPM are difficult to understand.

Despite increasing consciousness about the need to consider comprehensibility of process models, little research has been undertaken in order to improve and understand the relationships between modeling elements and comprehensibility. In this paper, we want to explore comprehensibility as a *local* property. This means that we measure the comprehensibility of a specific part of a BPM instead of the model as a whole. This way, we seek to investigate, which relations between elements in a graphical BPM are difficult to understand. In the research area of software complexity metrics, similar research has been published by Yang et al. [10]. Research results suggest that local complexity metrics could be a promising predictor for understandability. Therefore, we address the question as to when or under what circumstances similar relationships with local metrics will emerge in the context of BPM.

Our motivation is to complement the existing stream of work on improving the comprehensibility of BPM by examining local comprehensibility. In contrast to existing research such as [7, 8] we assign own metrics to each comprehension question in our study, not to the models as a whole.

The remainder of this paper proceeds as follows: First, comprehensibility of process model elements and their relationships is placed in context with a review of relevant theoretical perspectives. Next, we articulate a research model. Then, we discuss design and findings of an empirical study. The final section discusses the limitations of our work and presents the implications of our research.

2 Measuring the Cognitive Load of Process Models

2.1 Comprehensibility and Cognitive Load Theory

For defining the term *comprehensibility*, we adapt the definition for understanding of computer programs given by Biggerstaff [11] by relating it to the modeling context and replacing the word “program” by “BPM”:

“A person understands a BPM when they are able to explain the BPM, its structure, its behavior, its effects on its operational context, and its relationships to its application domain in terms that are qualitatively different from the tokens used to construct the BPM in a modeling language.”

Further popular explanations of the term *comprehensibility* such as “the ease with which the ... model can be understood” [12] suggest that cognitive effort is an important factor determining model comprehensibility and should be as low as possible. Based on the complex relationships and control flow logic of organizational processes in practice, understanding of BPM is a task likely to demand high cognitive effort.

For conceptualizing model comprehensibility in greater detail we draw on the notion that understanding of a fact in a BPM becomes more difficult if the number of model elements that need to be attended to increases. This is backed by the work on Cognitive Load Theory. Cognitive Load Theory builds on the fact that the capacity of the working memory at a given point of time is limited [13]. If the amount of information to be processed exceeds this capacity, comprehension is affected negatively. It has been shown that an instructional design that avoids an overload of the working memory makes understanding of the instructional material easier [14]. Prior research on various visual languages like entity-relationship models [15] or UML class diagrams [16] suggests that reducing the cognitive load improves the understandability of visual models.

2.2 Influence Factors for Model Comprehensibility

To determine the relevant factors for the cognitive load involved in understanding elements and their relations in a model, we draw on work on BPM metrics.

Relations between Elements Based on the similarity between structures in software code and process models, research results on code comprehensibility can serve as a profound basis for analyzing BPM comprehensibility. A large body of research exists on the cognitive complexity of different programming elements. Different control structures demand e.g. different levels of effort for understanding [17]. Little research has been undertaken to investigate the cognitive difficulty of different understanding tasks in process models. First efforts have been made by Melcher et al. [18]. In an experiment with 42 students reading a rather small BPM (containing 12 activities) they found that understandability values for questions on the four aspects order, concurrency, repetition and exclusiveness are different. As this is the only study in this context, further empirical research still needs to be done. Additionally, there is another strand of research stemming from the area of cognitive psychology, relating control flow elements and cognitive effort. Research on deductive reasoning has shown that systematic fallacies (so called ‘illusory inferences’) can occur when individuals construct or interpret mental models on premises concerning modeling-level connectives (like conjunctions or disjunctions) [19]. This situation may also be present for externalized visual BPM and may lead to higher error rates for understanding specific control flow elements. The current body of literature on error analysis of process models suggests for instance the existence of systematic reasoning fallacies concerning routing elements as inclusive OR gateways [20].

Element Interactivity The cognitive load that a task imposes on a person is represented by the number of elements that have to be attended to. This number is determined by the level of interactivity between the elements. Elements interact if they are interrelated such that it is necessary to assimilate them simultaneously [21]. High interactivity leads to high cognitive load because each element has to be processed with references to other elements. On the other

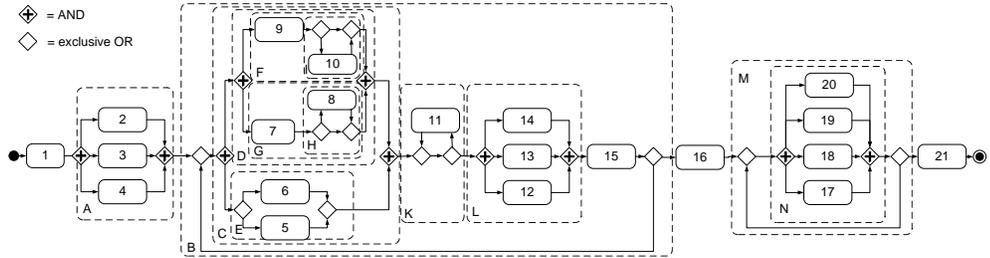


Fig. 1. business process model, structured into regions

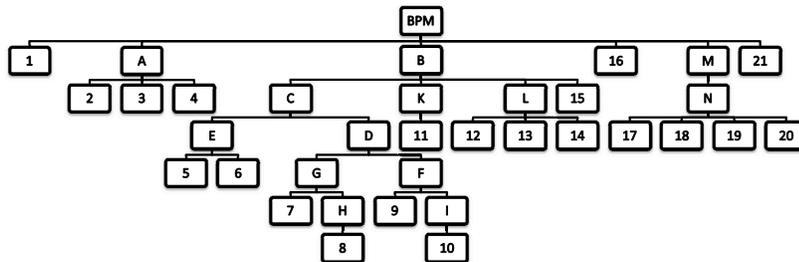


Fig. 2. PST for the model shown in Fig. 1

hand, cognitive load is low if the elements can be processed serially without referring to other elements.

In order to define a measure for the cognitive load resulting from the effort to understand the relation between two elements in a BPM, we follow the idea of Vanhatalo et al. [22] to decompose the BPM into canonical fragments with a single entry and a single exit. These fragments can be arranged in a process-structure tree (PST) such that there is exactly one PST for each BPM. For details we refer to [22], but we introduce the concept of a PST by an example. Fig. 1 shows a BPM (similar to the ones used in our experiment) and its canonical fragments that form the PST. Additionally to the fragments that are marked with dotted boxes, all single activities and the model as a whole are canonical fragments in the PST. From the example, it can be seen that canonical fragments can be nested. For example, the fragments D and E are within a larger fragment C. The depth of the nesting shows how many routing constructs in the BPM have to be understood in order to reason about the execution of an activity. The PST of the model is shown in Fig. 2. For a better readability, the control nodes (called gateways in BPMN) are omitted in this graph.

We argue that the distance between two elements in the PST can serve as a measure for the interactivity between those elements. Each region in the PST represents one concept (for example the concept of an exclusive choice or the concept of parallel branching) that the reader of the model has to understand. If elements are located in deeply nested control-flow blocks, the reader has to

understand a large number of concepts before being able to answer a question on the relation between those elements. In this case, the path between the two elements in the PST contains many arcs. On the other hand, if both elements are located into the same control block without additional nesting, they will also be in the same region of the PST, i.e. there are exactly two arcs in the PST between the elements. The assumption that the PST-distance can be an indicator of the difficulty to reason about a relation between two model elements is in line with the conceptual model of cognitive complexity by Cant et al. [23] that has been developed with respect to understanding software. Cant et al. discuss nesting within a a piece of software and argue that “the number of ‘steps’ [groups of control-flow statements; note from the authors] involved indicates the number of chunks which need to be considered” [23].

Formally, we define the *PST-distance* between two elements A and B of a BPM as the number of arcs between A and B in the PST minus one. This means that elements in a sequence or in the same control block have a PST-distance of 1. For example, in Fig. 1 the activities 17 and 18 which are executed in parallel inside the same control block have a PST-distance of 1 while the activities 16 and 17 (the latter is inside the fragments M and N) have a PST-distance of 3.

Element Separateness: Cut-Vertices A second aspect we take into account when discussing the interactivity between elements A and B in a BPM is the special case where a single arc in the BPM separates the BPM into two disjoint parts P_1 and P_2 such that $A \in P_1$ and $B \in P_2$.

In terms of graph theory this means that the connected graph G that forms the BPM has a so-called cut-vertex on a path from A to B , i.e. a vertex that when removed causes that the remaining graph is not connected anymore. If such a cut-vertex between A and B exists, the mental model of the relationships between A and B becomes much easier, because A is located “before” and B is located “after” an easy-to-spot reference point (the cut-vertex). For example, in Fig. 1 it is easy to see that activity 7 cannot be executed after activity 17. Because of the cut-vertices before and after activity 16, this can be concluded without analyzing the control structures in which the activities 7 and 17 are embedded. The assumption that the presence of a cut-vertex makes it easier to understand a model is backed by results by Mendling and Strembeck [2] who found that a large number of cut-vertices in a model has a positive effect on its understandability.

3 Research Model

Having laid out the relevant theoretical factors related to local understandability of process models, we will now draw several propositions to suggest how these factors will influence cognitive difficulty in comprehension tasks. Prior research on process model comprehension has almost exclusively focused on global model understanding, a focus of study that we extend in this paper by looking at the understandability of relations between elements in a process model.

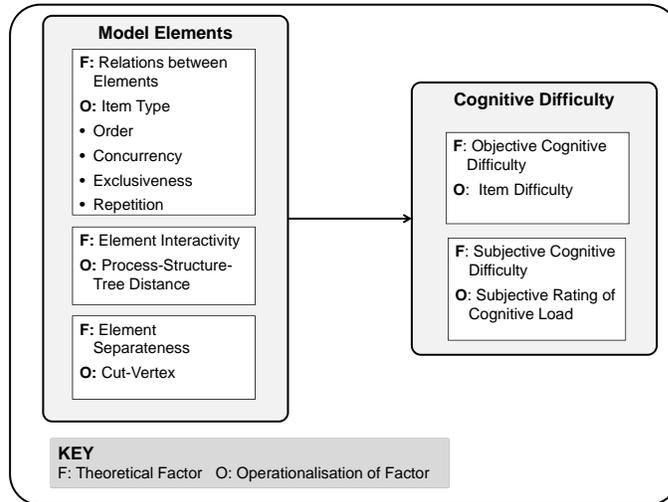


Fig. 3. Research Model

Fig. 3 shows our research model. The model proposes that the cognitive difficulty of understanding the relation between model elements is influenced by three factors: the type of relation between elements that has to be understood, the interactivity and the separateness of elements.

Following the research model, we now discuss three expected effects. As we anticipate similar effects on both objective as well as subjective side of the dependent variable 'cognitive difficulty', we formulate hypotheses for cognitive difficulty in general. First, we turn to different relations between elements. We state:

H1. The type of relation between elements that has to be understood (order, concurrency, repetition, exclusiveness) will have an influence on cognitive difficulty of understanding.

Second, we turn to the interactivity between elements. We expect that it is more difficult to understand relations between elements with a large PST-distance between them. Therefore, we have:

H 2. The interactivity between elements (high PST-distance) will be positively associated with the cognitive difficulty of understanding the relation between them.

Additionally we hypothesize if separateness of elements is low, understanding their relation gets easier:

H 3. High separateness between elements (existence of a cut-vertex between those elements) will be negatively associated with the cognitive difficulty of understanding the relations between them.

4 Research Method

4.1 Design and Measures

To test our hypotheses, we conducted an experiment in which the participants had to answer questions on a BPM. Model understandability (in terms of correctly answered questions) and perceived subjective difficulty were measured at each of the four levels *order*, *concurrency*, *repetition* and *exclusiveness* of the factor “type of comprehension question”. To manipulate the main factor we constructed comprehension questions targeting the four different relations between activities.

Comprehension Questions When selecting the questions, we took into consideration the work of Melcher et al. [18]. However, in comparison to [18] we formulated questions consistently, so that participants always had to consider two model elements (two activities) and their relationship for answering a question. Additionally we tried to use every-day-language in the questions. We used two different wordings to ask for the four relations between activities. To demonstrate the type of questions we refer to two activities with alphabetic names, although A and B were replaced with activity labels in the test material:

- Concurrency:
 - “A and B can be executed at the same point of time.”
 - “A and B can be executed in parallel.”
- Exclusiveness:
 - “In one process instance, A as well as B can be executed.”
 - “The activities A and B are mutually exclusive.”
- Order:
 - “If A as well as B are executed in a process instance, then A has to be finalized, before B can start.”
 - “If A as well as B are executed in a process instance, then A is executed before B.”
- Repetition:
 - “A can be executed more often than B.”
 - “In each process instance A is executed exactly as often as B.”

We took care that the wording in the questions is understandable, and we ran a pre-test in order to make sure that the participants understood the questions [24]. The comprehension questions, to which participants had to give a response of “right”, “wrong” or “I don’t know”, were selected so that each activity was addressed approximately once in each diagram. The response option “I don’t know” was included to lower guessing probability.

Questionnaire Construction For each model in the questionnaire we posed the same eight types of comprehension questions. Despite the use of the same wording, it is obvious that there is a large number of possibilities how to ask these questions, because any two activities can be targeted with the same question. We identified two basic variations: 1) the statement given in the question is correct or wrong and 2) the location of the chosen activities. For varying the

	Version A	Version B
	<ul style="list-style-type: none"> • Concurrency 1: L and M can be executed at the same point of time. (correct, close) 	<ul style="list-style-type: none"> • Concurrency 1: D and P can be executed at the same point of time. (wrong, distant)
	<ul style="list-style-type: none"> • Concurrency 2: G and S can be executed in parallel. (wrong, distant) 	<ul style="list-style-type: none"> • Concurrency 2: G and H can be executed in parallel. (wrong, close)
	<ul style="list-style-type: none"> • Exclusiveness 1: In one process instance E as well as F can be executed. (correct, close) 	<ul style="list-style-type: none"> • Exclusiveness 1: In one process instance K as well as S can be executed. (correct, distant)
	<ul style="list-style-type: none"> • Exclusiveness 2: The process steps C and R are mutually exclusive. (wrong, distant) 	<ul style="list-style-type: none"> • Exclusiveness 2: The process steps E and F are mutually exclusive. (wrong, close)
	<ul style="list-style-type: none"> • Sequence 1: If T as well as J are executed in a process instance, then T has to be finalized, before J can start. (wrong, distant) 	<ul style="list-style-type: none"> • Sequence 1: If A as well as C are executed in a process instance, then A has to be finalized, before C can start. (correct, close)
	<ul style="list-style-type: none"> • Sequence 2: If B as well as A are executed in a process instance, then B is executed before A. (wrong, close) 	<ul style="list-style-type: none"> • Sequence 2: If B as well as T are executed in a process instance, then B is executed before T. (correct, distant)
	<ul style="list-style-type: none"> • Repetition 1: U can be executed more often than D. (correct, distant) 	<ul style="list-style-type: none"> • Repetition 1: L can be executed more often than M. (wrong, close)
	<ul style="list-style-type: none"> • Repetition 2: In each process instance Q is executed exactly as often as P. (correct, close) 	<ul style="list-style-type: none"> • Repetition 2: In each process instance O is executed exactly as often as I. (correct, distant)

Fig. 4. Example of Asking Model Comprehensibility Questions

location of activities consistently, we decided to use pairs of activities, which are either close (≤ 1 activity between them) or distant (> 1 activity between them). As a consequence, we constructed the test material, such that each question was used once in each of four constellations (correct-close, correct-distant, wrong-close, wrong-distant), leading to 32 different question instances.

To ensure reliability of measurement we used a replication of the study design (questionnaire version A and B). In the replication, exactly the same models and comprehension questions were used, but the questions were asked for different activities in another constellation. Fig. 4 demonstrates how questions were asked for a specific process model.

Measured Variables The outcome of our main dependent variable *comprehension* is cognitive per se, i.e. it is created in the viewer's cognition and not directly observable. Therefore, it can only be measured indirectly or via comprehension questions. According to Aranda et al. [9] there are four variables that can measure comprehensibility: *correctness* (did the participant give the right answer?), *confidence* (certainty of the participant in his answers), *perceived difficulty* (to answer the question, as subjective judgment by the participant) and *time* (required to give an answer). In our experiment, we chose to use the main objective and subjective measure of cognitive difficulty, viz. the percentage of correct answers (correctness) as objective measure and the user's rating of cognitive load as subjective measure (perceived difficulty). To measure the perceived difficulty, we asked the users to rate it on a 7-point Likert-scale (with the labels "very difficult", "difficult", "rather difficult", "neither difficult nor easy", "rather easy", "easy" and "very easy").

4.2 Materials

Questionnaire Parts We used a pencil-and-paper questionnaire including three different sections in the experiment. The first section comprised items to obtain information about participants' demographic data, academic qualifications and modeling experience. Participants were asked about the number of years they had worked in the IT sector and the extent to which they had previously been involved with modeling in the context of education and work. After the first section, the questionnaire included a tutorial on process modeling, which covered all aspects the participants would need to know to perform the comprehension tasks. The third section included four different models with eight corresponding comprehension tasks per model. The amount of models used was determined by the selection of the comprehension questions during the experiment, as we wanted to ask 32 different instances of comprehension questions. To avoid order effects due to decreasing motivation or concentration of participants, we used two different scramblings. Models as well as comprehension questions were presented in different order, respectively.

Model Domain The four models were selected from different domains such that we could expect that they are understandable for an average student with no special domain knowledge.

Model Language Because it has been shown that the graphic style of a modeling language can influence the understandability of the model [25], we presented BPM modeled using different graphic styles. The models were modeled in different modeling directions and with three different routing symbol designs (UML Activity Diagrams, BPMN and Event-Driven Process Chains). These variations were included for allowing to generalize findings beyond specific layout and design restrictions. Additionally, they served as an experimental control to prevent a possible bias due to choosing only one modeling direction and routing symbols from a specific modeling language for all diagrams.

Model Layout We took into account that a change in the graphical layout of a BPM can influence its comprehensibility [9]. For this reason, we took care that the graphical layout of the models did not impose additional challenges to the reader.

Model Size Each of the four models used contained 21 activities. The model size was held constant for all models, because this variable is likely to have an influence on cognitive load of answering understandability questions.

4.3 Participants

A total of 199 business students participated in this study (125 males, 74 females), aged 23.5 years on average. Of all respondents, 36% were undergraduate students, 60% were master's level students and 4% had already completed their master's degrees. 67% had received training in modeling at university with 1.6 credit hours on average. About half of participants were familiar with Event-Driven Process Chains (60%) and UML Activity Diagrams (50%). 27% had work experience in the IT industry and 10% had already worked with BPMs.

5 Results

We first screened the data for its conformance with the assumptions of our statistical test. One assumption behind the use of ANCOVAs is that the variables are normally distributed. Kolmogorov-Smirnov tests confirmed that the dependent variables “percentage of correct answers” and “perceived difficulty” met this criterion ($p = 0.105$ and $p = 0.722$).

For each dependent variable, we ran a univariate ANCOVA with “relations between elements” and “element separateness” as independent factors and “element interactivity” as a covariate. According to the respective hypothesis, the percentage of correct answers and the perceived difficulty were the dependent variables. The ANCOVAs allow us to test the influence of the three predictor variables (two independent variables and a covariate) on the dependent variables as well as possible interaction effects. We use “element interactivity” as a covariate, because it is a continuous variable, which co-varies with the dependent variables percentage of correct answers ($r = -0.37$, $p = 0.002$) and perceived difficulty ($r = -0.61$, $p < 0.001$). Therefore, the covariate accounts for variance in the dependent variables and the inclusion of the covariate can increase the statistical power of the procedure.

5.1 Results for Hypothesis 1

ANCOVA results indicate that there is a effect of different types of relations between model elements (order, concurrency, repetition and exclusiveness) on cognitive difficulty of understanding their relation (H1). While there is only a trend for the percentage of correct answers ($F_{3,55} = 2.65$, $p = 0.058$), the effect on perceived difficulty is significant ($F_{3,55} = 4.20$, $p = 0.010$). According to our results, Hypothesis 1 is supported concerning subjective (perceived) difficulty, but only tentatively concerning objective difficulty (correct answers). Table 1 gives the percentage of correct answers and perceived difficulty for each type of relation. Order was the easiest relation (80% correct answers) with the lowest subjective difficulty (3.08), followed by concurrency (83%, 3.20). Exclusiveness was most the most difficult relation concerning correct answers (70%, 3.19) and repetition was rated as the most difficult by participants (71%, 3.58).

type of relation	percentage of correct answers		perceived difficulty	
	Mean	SD	Mean	SD
order	85%	6.16	3.08	0.35
concurrency	83%	10.09	3.20	0.28
repetition	71%	11.40	3.58	0.33
exclusiveness	70%	23.73	3.19	0.43

Table 1. Results for Hypothesis 1 (influence of types of relations between model elements), scale for perceived difficulty from 1=“very easy” to 7=“very difficult”)

PST-distance	cases	percentage of correct answers		perceived difficulty	
		Mean	SD	Mean	SD
1	19	80.2%	20.62	2.89	0.31
2	13	79.5%	10.35	3.30	0.26
3	15	76.9%	15.06	3.42	0.27
4	5	76.4%	76.38	3.47	0.54
5	6	64.9%	64.85	3.55	0.29
6	4	74.8%	74.80	3.47	0.36
7	2	79.6%	79.64	3.54	0.28

Table 2. Results for Hypothesis 2 (Influence of PST-Distance)

5.2 Results for Hypothesis 2

As expected, the covariate PST-distance (element interactivity) has an influence on the percentage of correct answers ($F_{1,55} = 4.32$, $p = 0.042$). Additionally there is a highly significant effect on perceived difficulty ($F_{1,55} = 22.04$, $p < 0.001$). Table 2 shows the average percentages of correct answers and the average perceived difficulties across different PST-distances. Hypothesis 2 predicted that PST-distance will be positively associated with cognitive difficulty. Therefore, Hypothesis 2 is supported.

5.3 Results for Hypothesis 3

79.9% of the questions about two activities with a cut-vertex between them have been answered correctly, compared to 75.8% of the questions about two activities without a cut-vertex. Although the difference between means shows in the expected direction, the results of the ANCOVA indicate that this difference is not statistically significant and that there is also no significant influence on perceived difficulty. Moreover, there are also no interaction effects of “element separateness” (cut vertex) with “relations between elements”.

Therefore, there is not enough evidence to support Hypothesis 3, which expected that the presence of a cut-vertex makes it easier to answer a question.

6 Discussion

This study provides empirical results on the influence of different relation types of elements, their interactivity and their separateness on cognitive difficulty of understanding the relation between elements.

In line with our predictions in Hypothesis 1, we found that different control structures in a BPM (like order or concurrency) differ according to their difficulty to be understood. Our results are in line with [18]. However, results are not directly comparable, as we used different wordings of possible understandability questions based on possible issues concerning ambiguousness (see [24] for details) and consistently addressed two model elements in the questions.

We further found that users perceive the relation between elements with a larger PST-distance as more difficult to understand. This effect has not been researched so far but is comparable to the discussion whether the nesting level in a BPM has an influence on its understandability. Mendling et al. [7, 8] did not find a significant relationship between the nesting level and the understandability of a model. However, while Mendling et al. regarded the nesting level as a global attribute of a BPM, we related the PST-distance to the model elements we asked about.

While our results on Hypothesis 2 support the theory that the PST-distance is correlated with the difficulty of a task, the results are still not yet conclusive.

In particular, in our experiment there were too few cases with a PST-distance greater than 3 to come to reliable results about deeply nested model elements. Furthermore, in the models we used for our experiment, the presence of a cut-vertex was more likely between elements with a large PST-distance (like the activities 10 and 20 in Fig. 1). This can explain the fact that understanding for the elements with PST-distance 6 and 7 was better than for those elements with PST-distance 5 (see Tab. 2). Future research on this question will be necessary.

An interesting observation was that in some cases, a small PST-distance can even mislead to a wrong conclusion. For the model shown in Fig. 1, we asked the question whether both activity 5 as activity 6 can occur in the same process instance. Because of the exclusive OR-gateway before those activities, 76% of the participants answered “no”. We assume that they did not bother to look at the parts of the model outside fragment E. Therefore they did not realize that this fragment is inside a loop and can be executed more than once.³

Our results did not confirm an influence of the existence of a cut vertex on the cognitive difficulty of an understanding task (H3). This relationship has also been discussed by other authors [7, 8, 2]. In contrast to our results, a similar experiment by Mendling and Strembeck [2] provided support for the hypothesis that a BPM with more cut-vertices is easier to be understood. Further studies [7, 8] yielded inconsistent results on this topic, so further research will be necessary.

From a more general perspective, our findings highlight that reducing the cognitive load improves the understandability of process models as already demonstrated for other visual models [15, 16]. Additionally our results provide support for the contention that the cognitive process for understanding a model depends on the actual task being carried out. This has already been substantiated in the research area of software comprehension by the work of Gilmore and Green [26].

7 Limitations

As with all studies, the research results presented here are subject to a number of limitations.

Model Size We acknowledge that our models might not be representative for all kinds of BPM. Models from real projects are often much larger than the

³ A very similar observation has been reported in [7].

ones used in our experiment. On the other hand, selecting rather simple models allowed us to keep the number of activities constant for all models and to avoid models that cannot be understood without the knowledge of a particular domain.

Questions Right/wrong questions can introduce a measurement error, because on average 50% of the questions will be answered by guessing alone. For this reason, we left the possibility of checking “I don’t know”. Additionally, we acknowledge that for some questions users might have guessed the expected answer based on domain knowledge. Future research should collect similar data sets based on models with meaningless activity labels like “activity XY” as suggested in [27]. While we did not find understanding problems during the pre-test, in the analysis we realized that the statement “The activities A and B are mutually exclusive.” gives room for misunderstandings (“A and B can not be processed either at the same time vs. both in the same process instance”). However, as those questions did not lead to more wrong answers as the alternative questions (see Sect. 4.1), we refrained from excluding these questions from our data.

Participants The participants of our study were students who were familiar with the modeling languages, although they were not experts in this area. The results might differ if the experiment is replicated with experts in business process modeling [28]. We tried to select participants for the experiment so that there was a variation of little to medium experience with conceptual modeling, resembling potential users in practice. However, the results might not be generalizable to the entire population of BPM users.

Selection of Influence Factors The factors for BPM understanding we have analyzed are not exhaustive. For example, we did not take into account the effect of the type of control structures (for example alternative or parallel branching) that are nested in the PST-tree. As related papers on this subject suggest that this factor should be considered as well [23], future research could examine this topic in detail.

8 Implications and Conclusion

This study is one of the first to investigate understandability as a local property in a model and denotes an important extension to the literature on influence factors for BPM understandability. Our main contribution is a first analysis of the cognitive difficulty of different relations between elements (order, concurrency, repetition, exclusiveness) in process models. Prior research has predominantly looked at global understandability of models and differences between models that influence understandability, in contrast we investigated local understandability of different items in a model.

Our results have implications for business process modeling practice and research. In terms of research, the results have an implication on the design of future experiments that measure understandability aspects of BPM. Our results demonstrate that several aspects of question selection (as the selection of the model elements and the type of the question) have an influence on cognitive difficulty. Implications of these results for researchers include exercising caution

when aggregating answering rates of randomly chosen comprehension questions to total comprehension measures for models. As the choice of questions might significantly influence comprehension scores, balanced selection and construction of questions is highly relevant.

In addition, our work provides further evidence that high interactivity of elements may heighten cognitive load and lower comprehensibility of BPM. If possible, deep nesting of control-flow blocks should be avoided in order to make understanding easier and – in the end – to improve the quality of BPM and reduce modeling errors. Research on modularity of BPM [1] suggests that decomposing complex models into smaller submodels can improve model comprehensibility. Additionally syntax highlighting [29] can be used to heighten comprehensibility of deeply nested blocks.

Future research is needed to determine valid and reliable values for the cognitive difficulty of understanding specific relations between model elements. These values could make it possible to finally estimate understandability of models without the need of a user evaluation. Looking ahead, exact comprehension values could then be used to guide modeling tool developers to provide feedback on cognitive difficulty of models to users or to give hints on possible understandability problems in models.

References

1. Reijers, H., Mendling, J.: Modularity in process models: Review and effects. In: Proc. of the 6th Int. Conf. on Business Process Management, Springer (2008) 20–35
2. Mendling, J., Strembeck, M.: Influence factors of understanding business process models. In: Business Information Systems. Volume 7 of LNBIP. Springer (2008) 142–153
3. Genon, N., Heymans, P., Moody, D., Amyot, D.: Improving the cognitive effectiveness of the bpmn 2.0 visual syntax. (2010)
4. Gruhn, V., Laue, R.: Complexity metrics for business process models. In: 9th International Conference on Business Information, Springer (2006) 1–12
5. Mendling, J.: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness. Volume 6 of LNBIP. Springer (2008)
6. González, L.S., Rubio, F.G., González, F.R., Velthuis, M.P.: Measurement in business processes: a systematic review. *Business Process Management Journal* **16** (2010) 114–134
7. Mendling, J., Reijers, H.A., Cardoso, J.: What makes process models understandable? In: Business Process Management, 5th International Conference. Volume 4714 of LNCS., Springer (2007) 48–63
8. Reijers, H., Mendling, J.: A study into the factors that influence the understandability of business process models. *IEEE Transactions on Systems, Man, and Cybernetics, Part A* (2010)
9. Aranda, J., Ernst, N., Horkoff, J., Easterbrook, S.: A framework for empirical evaluation of model comprehensibility. In: MISE '07: Proceedings of the International Workshop on Modeling in Software Engineering. (2007)

10. Yang, J., Hendrix, T.D., Chang, K.H., Umphress, D.: An empirical validation of complexity profile graph. In: Proceedings of the 43rd annual Southeast regional conference - Volume 1. ACM-SE 43, New York, NY, USA, ACM (2005) 143–149
11. Biggerstaff, T.J., Mitbander, B.G., Webster, D.: The concept assignment problem in program understanding. In: ICSE '93: Proceedings of the 15th international Conference on Software Engineering. (1993) 482–498
12. Moody, D.L.: Metrics for evaluating the quality of entity relationship models. In: Conceptual Modeling ER '98. Number 1507 in LNCS. Springer (1998) 211–225
13. Kirschner, P.A.: Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction* **12** (2002) 1–10
14. Sweller, J.: Cognitive load during problem solving: Effects on learning. *Cognitive Science* **12** (1988) 257–285
15. Moody, D.L.: Cognitive load effects on end user understanding of conceptual models: An experimental analysis. In: Proceedings of 8th East European Conference on Advances in Databases and Information Systems, ADBIS 2004. Number 3255 in LNCS, Springer (2004) 129–143
16. Genero, M., Manso, E., Visaggio, A., Canfora, G., Piattini, M.: Building measure-based prediction models for UML class diagram maintainability. *Empirical Software Engineering* **12** (2007) 517–549
17. Cant, S.N., Jeffery, D.R.: A conceptual model of cognitive complexity of elements of the programming process. *Information and Software Tech.* **37** (1995) 351–362
18. Melcher, J., Mendling, J., Reijers, H.A., Seese, D.: On measuring the understandability of process models. In: Revised Papers of the BPM 2009 International Workshops. Volume 43 of LNBIP., Ulm, Germany, Springer (2010) 465–476
19. Khemlani, S., Johnson-Laird, P.N.: Disjunctive illusory inferences and how to eliminate them. *Memory & Cognition* **37** (2009) 615–623
20. Mendling, J., Neumann, G., van der Aalst, W.M.P.: Understanding the occurrence of errors in process models based on metrics. In: On the Move to Meaningful Internet Systems 2007. Volume 4803 of LNCS. Springer (2007) 113–130
21. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* **4** (1994) 295 – 312
22. Vanhatalo, J., Völzer, H., Koehler, J.: The refined process structure tree. *Data & Knowledge Engineering* **68** (2009) 793 – 818
23. Cant, S., Jeffery, D., Henderson-Sellers, B.: A conceptual model of cognitive complexity of elements of the programming process. *Information and Software Tech.* **37** (1995) 351 – 362
24. Laue, R., Gaddatsch, A.: Measuring the understandability of business process models - are we asking the right questions? In: 6th International Workshop on Business Process Design. (2010)
25. Nordbotten, J.C., Crosby, M.E.: The effect of graphic style on data model interpretation. *Inf. Syst. J.* **9** (1999) 139–156
26. Gilmore, D.J., Green, T.R.G.: Comprehension and recall of miniature programs. *International Journal of Man-Machine Studies* **21** (1984) 31–48
27. Parsons, J., Cole, L.: What do the pictures mean?: guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. *Data Knowl. Eng.* **55** (2005) 327–342
28. Petre, M.: Why looking isn't always seeing: readership skills and graphical programming. *Commun. ACM* **38** (1995) 33–44
29. Reijers, H., Freytag, T., Mendling, J., Eckleder, A.: Syntax highlighting in business process models. *Decision Support Systems to appear* (2011)