Exploring the Cognitive Effects of Ambiguity in Process Models

Marco Franceschetti[1(✉)], Amine Abbad-Andaloussi[1], Clemens Schreiber[2], Hugo A. López[3], and Barbara Weber[1]

[1] Institute of Computer Science, University of St.Gallen, St.Gallen, Switzerland
{marco.franceschetti@unisg.ch}
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
[3] Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract. Ambiguity in business process models might lead to multiple alternative process interpretations by the readers. This plurality of interpretations causes undesirable situations such as misunderstandings, unclear responsibilities, and unexpected behaviors. However, to date, little attention has been given to how ambiguity affects the model readers. Here, we report on an eye-tracking study aimed at investigating the impact of different ambiguities (i.e., pragmatic, semantic, syntactic, and lexical) on readers' cognitive load, comprehension, and visual associations when reading process models. The results of this study show that these ambiguities yield a significant impact on cognitive load, comprehension, and visual associations. These results raise further attention toward the negative effects of ambiguity from a cognitive and behavioral perspective, and stimulate the development of novel tools supporting ambiguity detection in process models.

Keywords: Ambiguity · Business process models · Eye-tracking.

1 Introduction

Business process models promise to enhance process understanding through visualization and formal modeling language semantics. However, despite the availability of formal semantics for modeling languages such as the Business Process Model and Notation (BPMN) [28], there is a growing recognition that business process models may be susceptible to ambiguity [11,12,34]. Ambiguity is a characteristic of a process model that can lead to multiple possible process interpretations. Recently, a comprehensive study of ambiguity in Business Process Management (BPM) highlighted possible sources of ambiguity in business processes [15]. Drawing from concrete examples in the literature and public datasets, the study underpinned the pervasive nature of ambiguity in process models.

The possibility that a process model may be interpreted in multiple different ways entails the risk of misunderstandings, unclear responsibilities, unexpected behaviors, and potential cascading effects across other process-related artifacts managed in the BPM lifecycle, as detailed in [15]. Ultimately, this compromises the benefits of adopting Process-aware Information Systems [21,29]. Indeed, a correct comprehension of the involved process models is imperative for the successful engineering and operation of a Process-aware Information System [29].

Prior works have studied different ambiguities in business process models, aiming at defining strategies to deal with or diminish ambiguity (cf. [12,26,34]). However, to the best of our knowledge, no prior work comprehensively studied ambiguity in relation to its impact on a model reader's cognitive and behavioral dimensions. While prior works investigated the impact on these dimensions with respect to aspects such as model complexity [33], model quality [19], modeling language notational deficiencies [14], or personal aspects of readers such as modeling experience and process knowledge [27], the impact of ambiguity remains, to date, unexplored. What sets ambiguity apart from these, and other aspects such as model correctness, is that ambiguity leads to a multiplicity of correct admissible interpretations. This is, in several processes such as normative ones, an intentional model feature to enable flexibility in the model interpretation [15]. How the presence of this specific characteristic, which can lead to multiple interpretations, impacts a model reader's cognitive and behavioral dimensions (for instance, by challenging the model comprehension) is the focus of this study.

In this work, we take on the challenge of *measuring the impact* of ambiguity in process models on model readers' cognitive load (i.e., mental effort), model comprehension, and visual associations (i.e., the shift of attention between model elements, which suggests the raise of mental demand to integrate different information, cf. [8]) using eye-tracking. We chose eye-tracking as it has been previously used to study process model comprehension tasks, providing significant insights in relation to cognitive dimensions (cf. [8,31,32,39,42]).

Our study is driven by the following research questions:

– RQ1. How do different ambiguities affect model readers' cognitive load?

– RQ2. How do different ambiguities affect model readers' comprehension of process models?

– RQ3. How do different ambiguities affect model readers' visual associations?

To answer RQ1–RQ3, we observed the eye movements of model readers while they performed different comprehension tasks on process models with and without ambiguities. Our results demonstrate the usefulness of eye-tracking to detect ambiguity, and that ambiguous process models lead to higher cognitive load (RQ1), challenged comprehension (RQ2), and increased visual associations (RQ3) when compared to process models without ambiguities. Having demonstrated the negative effects of ambiguity, future work could focus on developing methods and techniques to support model readers by automatically detecting ambiguities in process models. Our results also stimulate further studies on the impact of ambiguity in relation to expertise and on disambiguation strategies.

This paper is structured as follows: in Section 2, we recall background concepts and formally define ambiguities in process models. In Sect. 3, we report on our experiment. In Sect. 4, we report on the findings. In Sect. 5, we discuss implications of the experiment results. Sect. 6 concludes the paper.

2 Background and Related Work

In this section, we set the theoretical underpinning on ambiguity in BPM and formally define four types of ambiguities in process models. We then set the theoretical underpinning on the cognitive theories relevant to our empirical study.

2.1 Ambiguity in Process Models

Ambiguity in BPM refers to the possibility that a business process admits multiple different interpretations. As per the characterization in [15], intrinsic (i.e., language-specific) or extrinsic (i.e., modeling task-specific) factors can induce ambiguity in process models. Here, we extend this characterization defining pragmatic, semantic, syntactic, and lexical ambiguities in process models, whose effects we investigate in this study. These definitions adapt and unify those found in existing literature that has dealt with these ambiguities (cf. [7,12,18,23,26,34]) or with quality issues leading to these ambiguities (cf. [22]). As detailed in Sect. 3, we deliberately introduced these ambiguities in the process models used in our study to assess their effect on model readers.

Definition 1 (Pragmatic Ambiguity). *(Adapted from [7,18]) Pragmatic ambiguity is a phenomenon that occurs at the layout level causing a process model to lack clarity in one or more process perspectives, allowing for multiple interpretations without affecting the behavior of the executable process model.*

An example of pragmatic ambiguity is shown in Fig. 1. Here, the overlapping control flow edges between the activities allow for multiple interpretations of the possible precedence constraints. The idea that presentation layout may cause ambiguity was introduced in [30]. Further studies such as [13,18,32] focused on the comprehension of process models in relation to layout.

Definition 2 (Semantic Ambiguity). *(Adapted from [12,18,22]) Semantic ambiguity is a linguistic phenomenon related to the usage of a modeling language that might occur when a process model lacks validity or completeness, allowing for multiple interpretations. A model is valid if all statements in the model are correct and related to the process. A model is complete if there is a one-to-one mapping between model constructs and domain concepts.*

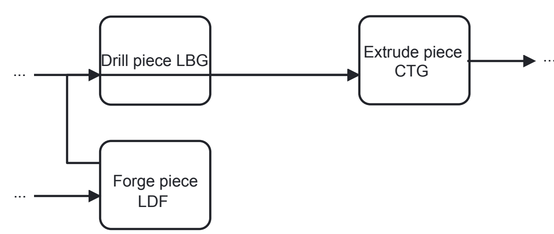

Fig. 1: BPMN fragment of a process with pragmatic ambiguity

An example of (validity-related) semantic ambiguity is shown in Fig. 2 (left), with the sequence *"Approve piece LDF"* – *"Reject piece LDF"*. Here, common sense suggests that a piece is either approved or rejected, and it is unclear whether the activities refer to different objects or they are not supposed to be both executed in the same trace. Semantic ambiguity of process models was investigated with a focus on BPMN models in [11]. Semantic issues in relation to process model quality were studied in [18,22]. In [12], authors proposed an ontology-based approach to reduce semantic ambiguity in process models.

Definition 3 (Syntactic Ambiguity). *(Adapted from [7]) Syntactic ambiguity is a grammatical phenomenon that occurs when a fragment of a process model M formalized in a modeling language $\mathcal{L}$ can be parsed using more than one grammatical structure of $\mathcal{L}$, allowing for multiple possible interpretations of M.*

An example of syntactic ambiguity is shown in Fig. 2 (right), in which the control flow splits at a XOR-gateway with no conditions attached. Here, the split can be interpreted either as an underspecified XOR-gateway (the following activities are mutually exclusive, although the condition is unknown) or as an AND-gateway (the following activities must be both executed) that has been assigned the wrong type. To promote the design of understandable and unambiguous process models, the work in [10] proposed a set of guidelines, among which several refer to syntactic aspects of process models. Syntactic issues leading to multiple interpretations of BPMN models were discussed in [18]. The work in [7] discussed syntactic ambiguity in relation to process descriptions.

Definition 4 (Lexical Ambiguity). *(Adapted from [34]) Lexical ambiguity is a linguistic phenomenon related to the usage of the natural language that occurs when a textual label in a process model can be interpreted in multiple ways.*

Abbreviations, homonyms, synonyms, and polysemic words (i.e, words with multiple meanings) are possible causes for lexical ambiguity [10,35]. For example, consider an activity labeled *"Receive report"* followed by an activity labeled *"Evaluate summary"*. Here, it is unclear whether the terms "report" and "summary" are used as synonyms and refer to the same data object, or they refer to different data objects. The use of certain labeling styles (e.g., passive voice, noun

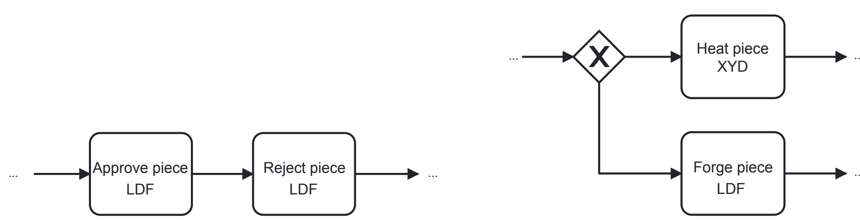

Fig. 2: BPMN fragments of processes with semantic (left) and syntactic (right) ambiguities

form of verbs) might as well lead to lexical ambiguity [26]. For an example borrowed from [34], consider an activity labeled *"Plan integration"*: here, one could interpret the activity either as the planning of some integration, or as the integration of some plan. Prior work from Mendling et al. investigated the usage of activity labels and the resulting lexical ambiguity [24,26]. These studies resulted in guidelines [25], refactoring recommendations [23], and automatic approaches to detect and resolve lexical ambiguity in process models [34].

2.2 Cognitive load, Comprehension, and Visual Associations

Here, we provide background and set the theoretical underpinnings for our empirical study, which investigates the impact of different ambiguities in process models on readers' *cognitive load*, *comprehension*, and *visual associations.*

Cognitive Load. The Cognitive Load Theory defines cognitive load as the amount of workload imposed on the human working memory during tasks requiring mental processing [37]. It arises from different sources, such as the inherent complexity of the material a reader engages with (leading to intrinsic load), the material representation (leading to extraneous load), and a reader's ability to integrate the material with their existing mental schemes (leading to germane load).

Ambiguous process models allow for a plurality of interpretations. Therefore, when attempting to resolve ambiguities in understanding process models, readers are expected to actively search for additional clues and information to ensure the most accurate interpretation. This inclination to seek more context than they would with unambiguous models is expected to elevate the necessity for augmented information processing, thereby increasing their cognitive load. This proposition finds support in Campbell's work [9]. In his review of existing literature, Campbell identified a set of features characterizing complex tasks that place high mental demands on the reader (i.e., cognitive load). One such feature is the existence of uncertain or conflicting information in the artifact. Another is the presence of multiple ways to solve the task, each with a similar level of viability. This, in turn, increases readers' cognitive load as they have to seek out additional information, evaluate, and choose among comparable alternatives.

Several measures have been used in the literature to estimate readers' cognitive load [3,13,40]. These measures can be organized into *subjective* and *objective.* Typically, the subjective measures are based on the reader's self-assessment of perceived difficulty [3]. This information can for instance be collected through a 5-points Likert scale (from 0: "very easy" to 4: "very difficult") asking the readers to rate the difficulty they perceived while solving a task (cf. [3,36,44]).

Beside subjective measures, objective measures provide more powerful means to assess readers' cognitive load [40]. Notably, through eye-tracking, one can investigate the characteristics of readers' fixations (i.e., the time the eye remains stationary at a specific coordinate of the stimulus, e.g., a process model [20]) to estimate their cognitive load objectively [20]. Using fixation features as an indicator of cognitive load takes root from the eye-mind hypothesis [20], postulating that the mind processes what the eyes are currently looking at. Glöckner and Herbold [16] advanced this theory by posting that fixations with a duration $\geq 250ms$ reflect mental processing, which can be associated with cognitive load.

To test the impact of ambiguities on cognitive load, we use both subjective and objective measures. For the former, we rely on a 5-points Likert scale questionnaire of perceived difficulty. As for the latter, we use the mean number of fixations $\geq 250ms$ as an indicator of cognitive load. We expect both measures to increase significantly when readers are dealing with ambiguous process models due to the mental processing needed when attempting to resolve the ambiguities.

Comprehension. The comprehension of process models has been subject to extensive research (cf. overview in [13]). According to [38], when readers experience very high levels of cognitive load, their performance is likely to decrease. Performance in comprehension tasks is usually captured through *comprehension accuracy* and *comprehension efficiency* [13,36]. The former is computed by comparing the outcome of the reader task (e.g., answer to a comprehension question) to a known ground truth (e.g., whether the answer is true or false); the latter is computed by calculating the time interval between the task start and end.

In our study, we use *comprehension efficiency* as a comprehension measure, which we expect to be reduced when dealing with ambiguous process models. We refrain from using *comprehension accuracy* as it is not meaningful to evaluate the correctness of comprehension questions on ambiguous process models due to the multiplicity of possible interpretations.

Visual Associations. Eye-tracking is a powerful tool for studying readers' visual behavior [20], especially when analyzing visual associations reflecting a repeated shift of attention between different model elements [8]. This behavior is typically interpreted as an attempt to mentally integrate visually associated elements [8]. Hence, through visual associations observed with eye-tracking, one can tap into readers' cognitive integration processes and study their characteristics [8]. When attempting to disambiguate a process model, the seeking of additional information spread over different parts of the model coupled with the constant need to consolidate this information is likely to manifest in a high frequency of visual associations. Following [8], this pattern suggests an intensified requirement for cognitive integration. This behavior aligns with Campbell's view of how users engage with ambiguous tasks where multiple potential paths to a resolution exist or the tasks comprise conflicting information [9].

Visual associations can be quantified using the AOI (Area of Interest) Run Count (AOIRC) [8]. This measure has been prominently deployed to capture visual associations and was proven to reflect readers' cognitive integration processes [8]. An AOI denotes a part of the stimulus with information relevant for a particular analysis [20]. When analyzing eye-tracking data on process models, an AOI refers to a model element (e.g., activity or gateway) [20]. The AOIRC quantifies the number of entries and exits to an AOI from other AOIs. As reported in [8], higher cognitive integration manifests in increased AOIRC. Visual associations can also be observed by discovering process maps reflecting readers' visual behavior [4,17]. With low visual associations, the process maps are expected to reflect a straightforward, sequential reading pattern; with high visual associations, they are expected to reveal a rather convoluted, non-linear reading pattern, marked by frequent transitions between the fixated model elements.

We use the AOIRC and process maps to study readers' visual associations and cognitive integration. We expect the AOIRC to increase when reading ambiguous process models due to the multiple interpretations they allow and the complexity of integrating such information in the readers' mental schemes. This effect would manifest and be observable in more complex and convoluted process maps.

3 Research Method

To investigate the impact of ambiguity on the cognition and behavior of model readers, we designed an eye-tracking study following the guidelines in [1]. Here, we report on the study design, execution, and subsequent data analysis.

3.1 Study Design

Research Model. The goal of our study is to investigate the impact of different ambiguity types in process models on model readers' cognitive load, comprehension and visual associations. To this end, we use a within-subject design following the research model depicted in Figure 3. Each ambiguity type (i.e., *pragmatic, semantic, syntactic* or *lexical*), denoting a theoretical construct on the independent variables side, is investigated separately with respect to its impact on *cognitive load*, *comprehension*, and *visual associations*, denoting the theoretical constructs on the dependent variables side.

At the independent variables side, each *type of ambiguity* (i.e., a factor) comprises two factor levels: *no ambiguity* and *ambiguity.* Accordingly, these factor levels are operationalized across a process model with 2 sub-processes *without* ambiguities, and one sub-process *with* an ambiguity of that specific type. At the dependent variables side, *cognitive load* is operationalized through *users' self assessment of perceived difficulty* and the *mean number of fixations associated with mental processing* (i.e., having a duration $\geq 250ms$) (cf. Sect. 2.2). *Comprehension* is operationalized using *comprehension efficiency* (i.e., time required for a task, cf. Sect. 2.2) and *visual associations* are operationalized with the *AOIRC measure* (cf. Sect. 2.2). All these measures are collected at the sub-process level.

Our within-subject design allows us to conduct a pairwise within-subject comparison between the two factor levels of each independent variable with respect to our dependent variables. Following the theoretical underpinnings presented in Sect. 2.2 on the effects of ambiguity on readers' cognitive load, comprehension, and visual associations, we formulate the following hypotheses **H1**–**H3**, which apply to each of the four ambiguities (when considering a specific ambiguity, we use the shorthands $\mathbf{Hx}_{prag}$, $\mathbf{Hx}_{sem}$, $\mathbf{Hx}_{syn}$, and $\mathbf{Hx}_{lex}$):

– **H1:** A sub-process model with ambiguity yields a higher cognitive load compared to a model without ambiguity.

– **H2:** A sub-process model with ambiguity yields lower comprehension compared to a model without ambiguity.

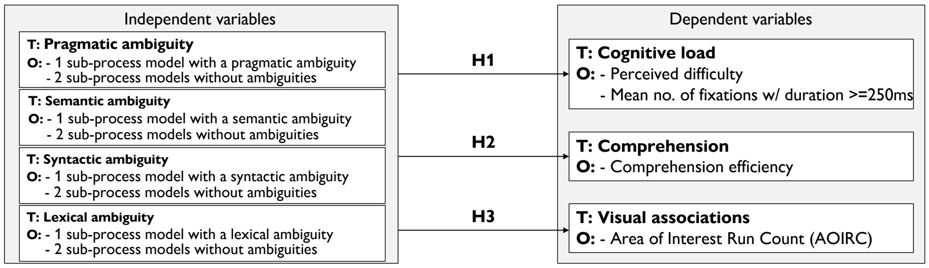

Fig. 3: Research model. T = theoretical construct; O = operationalization of the construct.

– **H3:** A sub-process model with ambiguity yields higher visual associations than a model without ambiguity.

Material. For the study material we designed 13 process models representing manufacturing processes using Camunda Modeler[4]. Each model included one ambiguity; no two models shared the same ambiguity. One model was meant to let the participants familiarize with the task and ask for clarifications, 3 models were affected by one pragmatic ambiguity each, 3 by one semantic ambiguity each, 3 by one syntactic ambiguity each, and 3 by one lexical ambiguity each. As several modeling guidelines have been proposed to avoid ambiguity in process models (cf. [10,25]), we designed each of the models deliberately violating specific guidelines with respect to a specific model element, e.g., a gateway or an activity label. These violations resulted in the model element being affected by ambiguity. For instance, in violation to guideline *35: Labeling AND-gateways* from [10], we added labels to an AND-split gateway, making the gateway ambiguous. Since the violated guidelines derive from multiple empirical studies [10,25], we consider this approach adequate to generate noticeable ambiguity by the study participants.

In line with our within-subject design, each model consisted of a sequence of 3 sub-processes. Ambiguity was present in only one of the 3 sub-processes, to be able to compare the cognitive and behavioral measures associated with the analysis of sub-processes with and without ambiguity. To mitigate confounding factors associated with model size and complexity, all sub-processes had similar size and complexity. They all had 10 activities, 1 intermediate event, and 3 pairs of split-join gateways; they all had similar complexity according to 17 complexity metrics. As domain knowledge may be a confounding factor that individual participants could possess and leverage to resolve the ambiguities [43], activity labels were randomly assigned to mitigate any such knowledge. We used a verb-object labeling style, with a random verb from the manufacturing domain followed by *"piece"* and a random string of 3 letters (e.g., *Extrude piece TRS*).

We designed one task per process model. Each task consisted of a question specific to the ambiguity affecting the model. For all tasks involving pragmatic,

[4] See `https://camunda.com/download/modeler/`

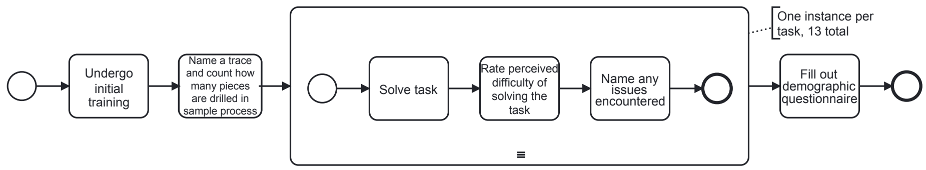

Fig. 4: Experiment procedure

semantic, and syntactic ambiguity, the question was *“Name a valid trace for the process, assuming that the temperature T=200”*. The motivation was that the question required the participants to sequentially analyze the control flow of the entire model (providing a specific temperature value (i.e., *T=200*) had the purpose of enforcing one specific trace). As pragmatic, semantic, and syntactic ambiguities affect the control flow, our specific task design ensured that participants have to face the associated ambiguity and interpret it to be able to solve the task. For instance, one model used for this task had the control flow edges after an AND-split gateway annotated with conditions: to answer the question, participants had to read the gateway and the associated conditions, and interpret them. In turn, for tasks involving lexical ambiguity, the control flow is not relevant; thus, we chose questions that rather emphasize the lexicon of the labels and asked questions such as *“How many pieces are moved?”* or *“How many pieces are tested?”*. The motivation was that these questions required the participants to carefully read all the activity labels in the model, including ambiguous labels, and to interpret them. For instance, for question *“How many pieces are tested?”*, the model had an activity labeled *“Design test of piece TRS”*, which has multiple interpretations (i.e., designing the test or testing the design of a piece) leading to different answers.

Participants. A total of 44 participants from St.Gallen University (20 participants) and Karlsruhe Institute of Technology (24) took part in the experiment. The age groups represented were 20–30, 30–40, 40–50, and 50+, with the majority (64%) of the participants falling into the 20–30 age group. Participants included 17 Bachelor’s and Master’s students, 18 PhD-level researchers, 4 IT employees, and 5 students also working as IT employees, with BPM expertise ranging from 0 to 25 years; they all had a basic understanding of BPMN or similar notations. Due to the within-subject experiment design, the participants’ heterogeneity does not affect the testing of our hypotheses.

3.2 Experiment Procedure

The experiment was conducted in individual sessions of around one hour in a controlled lab environment. Figure 4 depicts a BPMN diagram representing the experiment procedure. First, participants were given initial training on the basic BPMN elements found in the models used in the study. To avoid any bias, they were not informed about the study goals and the topic of ambiguity in process models. Then, they were asked to try to name a valid trace and count how many pieces are manipulated in an example process to ensure that they understand the kind of questions used in the experiment (cf. Sect. 3.1). During this phase, they could receive clarifications and feedback. Then, the actual experiment started with a familiarization task to let the participants familiarize themselves with the interface and tasks. To avoid any bias due to unfamiliarity with the experiment setup, the results of this task were not used for the subsequent data analysis. The presentation order of the next tasks was randomized to mitigate potential learning and fatigue effects (allowing them to spread uniformly across the different experiment tasks). After each task, participants rated on a Likert scale the perceived difficulty of solving the task with respect to each sub-process. This was used to compute the perceived difficulty measure of cognitive load (cf. Sect. 3.1). They were also generically asked to state whether they had any issues in solving the task and how they overcame them. This was used to check whether they had noticed the ambiguity in the process model without hinting at the presence of the ambiguity in the process model. We used this information in the data analysis to consider only tasks where ambiguity was noticed (cf. Sect. 3.3).

3.3 Data Collection and Analysis

The data was collected using EyeMind [5]–an eye-tracking tool for capturing eye-tracking data on process models within an interactive editor allowing users to seemingly browse the different parts of the model and explore its sub-processes. The key advantage of EyeMind lies in its support of dynamic eye-tracking stimuli where users can freely browse different views, scroll, and zoom in different parts of the stimulus [5,20]. Conducting experiments on a dynamic stimulus is known to be complex and time-consuming [5,20]. Hence, researchers typically adopt a static stimulus using a small and non-interactive process model (shown as an image), not reflecting the true complexity and usability of real-world process models, thus limiting the ecological validity (i.e., ability to generalize) of the insights [5,20]. To overcome the limitations of using a static stimulus and align with real-world model size and complexity, we opted for EyeMind, which, to the best of our knowledge, is the only tool supporting this feature on process models.

Following the data collection, we selected the trials [5] where the participants had correctly noticed the ambiguities embedded in the models, based on the premise that unnoticed ambiguities will presumably have no effect on cognitive load, comprehension and visual associations. From 528 trials (due to 44 participants performing 12 tasks, excluding the familiarization task), ambiguities were noticed in 386 trials (i.e., 73%). Then, we calculated the measures capturing the constructs depicted at the dependent variables side of our research model (cf. Fig. 3). These calculations were conducted at the level of each sub-process.

Participants were given 3 tasks per ambiguity type (cf. Sect. 3.1), hence we collected 3 data points per factor level and participant. To avoid interdependencies between the data points, for each participant we computed the

[5] A trial refers to an instance of a participant performing a task

mean value of each measure at each factor level. We computed the descriptive and inferential statistics reported in Table 1. The descriptive statistics allow for a *pairwise* within-subject comparison between the mean values of each measure across the factor levels: *presence* and *absence* of each ambiguity type. To ascertain the statistical significance of the differences between the means, we used the Wilcoxon Signed-Rank inferential test [41], as it is adequate for pairwise within-subject comparisons and does not make assumptions on the normal distribution of the data.

3.4 Data Availability and Reproducibility

In the spirit of transparency, reproducibility, and replicability, we published a permanently accessible online appendix, which includes:

- the complete set of process models used in the study;
- a detailed report on the process models' complexity metrics and how they have been computed;
- the specific guidelines violations and the resulting ambiguities;
- the list of tasks performed by the study participants;
- the participants demographics;
- a higher resolution version of Fig. 4 depicting the experiment procedure;
- the Python notebooks used for the data analysis and the analysis results.

The appendix is accessible at `https://zenodo.org/records/10715134`.

4 Findings

Here, we summarize our findings with respect to the research questions.
RQ1. How do different ambiguities affect model readers' cognitive load? We measure cognitive load in terms of perceived difficulty and mean number of fixations with duration $\geq 250ms$, collected at the level of each ambiguous and non-ambiguous sub-process model (cf. Sect. 2.2). As Table 1 shows, in all tasks ambiguous sub-process models yield higher perceived difficulty compared to non-ambiguous sub-process models. The reported p-values indicate that these differences are statistically significant. For the mean number of fixations $\geq 250ms$, significant differences can be observed for pragmatic, semantic, and syntactic ambiguity. In turn, for tasks dealing with lexical ambiguity, with a p-value of 0.634 no significant differences could be observed. These results confirm hypotheses $\mathbf{H1}_{prag}$, $\mathbf{H1}_{sem}$, and $\mathbf{H1}_{syn}$ and reject $\mathbf{H1}_{lex}$.
RQ2. How do different ambiguities affect model readers' comprehension of process models? We measure comprehension efficiency in terms of the time required to complete a task (cf. Sect. 2.2). Table 1 reports such mean times with respect to non-ambiguous and ambiguous sub-process models. For all tasks, the time required for ambiguous sub-process models was higher than for non-ambiguous models. As the p-values indicate, these differences were statistically significant, confirming hypotheses $\mathbf{H2}_{prag}$, $\mathbf{H2}_{sem}$, $\mathbf{H2}_{syn}$, and $\mathbf{H2}_{lex}$.

Table 1: Descriptive and inferential statistics. Comprehension Efficiency unit: milliseconds. *Note:* p <0.05 informs that the pairwise difference of means between the no ambiguity and ambiguity levels is significant.

A.	H.	Measure	Descriptive		Inferential
			No Ambiguity	Ambiguity	
			Mean	Mean	p-value
Pragmatic A.	H1	**Cognitive Load**			
		Perceived Difficulty	1.038	**3.331**	**<.001**
		Fixations >= 250 ms	32.754	**74.508**	**<.001**
	H2	**Comprehension**			
		Comprehension Efficiency	27198.497	**51667.651**	**<.001**
	H3	**Visual Associations**			
		AOI Run Count	21.159	**46.357**	**<.001**
Semantic A.	H1	**Cognitive Load**			
		Perceived Difficulty	0.893	**2.240**	**<.001**
		Fixations >= 250 ms	28.354	**41.547**	**<.001**
	H2	**Comprehension**			
		Comprehension Efficiency	25809.034	**39514.319**	**<.001**
	H3	**Visual Associations**			
		AOI Run Count	20.067	**34.573**	**<.001**
Syntactic A.	H1	**Cognitive Load**			
		Perceived Difficulty	0.973	**2.217**	**<.001**
		Fixations >= 250 ms	25.212	**34.308**	**<.001**
	H2	**Comprehension**			
		Comprehension Efficiency	26414.001	**32977.135**	**<.001**
	H3	**Visual Associations**			
		AOI Run Count	19.151	**27.489**	**<.001**
Lexical A.	H1	**Cognitive Load**			
		Perceived Difficulty	0.786	**1.976**	**<.001**
		Fixations >= 250 ms	**21.610**	21.463	0.634
	H2	**Comprehension**			
		Comprehension Efficiency	16258.680	**19436.565**	**0.019**
	H3	**Visual Associations**			
		AOI Run Count	18.086	**22.610**	**0.008**

RQ3. How do different ambiguities affect model readers' visual associations? To measure such impact of ambiguities, we measure the AOIRC (cf. Sect. 2.2). As Table 1 reports, in all tasks, the AOIRC for ambiguous sub-process models was higher compared to non-ambiguous sub-process models. All these differences were statistically significant in terms of p-values, confirming hypotheses $\mathbf{H3}_{prag}$, $\mathbf{H3}_{sem}$, $\mathbf{H3}_{syn}$, and $\mathbf{H3}_{lex}$. The process maps in Figure 5 complement

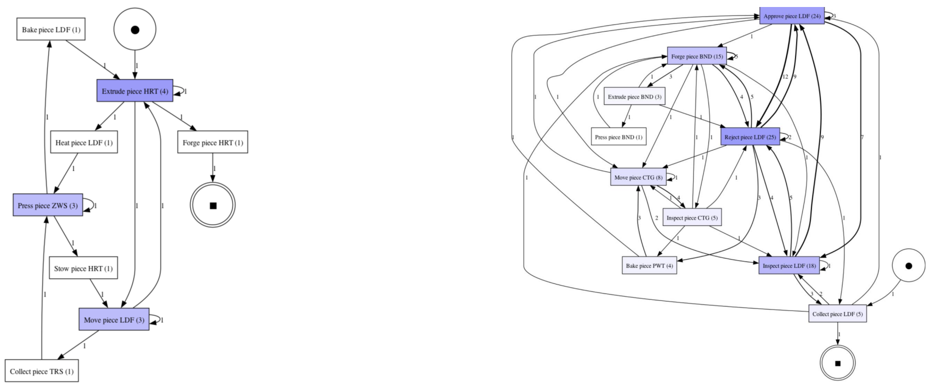

Fig. 5: Process maps comparing the visual associations of a participant when reading a sub-process without (left) and with (right) a semantic ambiguity. A higher resolution of this figure is available in the online appendix. The circle with a dot inside denotes the process start, the double circle with a square inside denotes the process end. Rectangles refer to visits to the different process model activities; edges refer to the transitions for visiting one activity from another. The color scale of the rectangles refers to the absolute visit frequency to an activity; the thickness and labels on the edges refer to the absolute transition frequency, resp. the number of transitions between each pair of activities.

the inferential statistics by showing an example of the striking differences in the visual associations (the shifts of attention between model elements) of a reader when analyzing a non-ambiguous (few shifts of attention) and an ambiguous (many shifts of attention) sub-process model within the same task. Analogous process maps for all tasks and participants are available in the online appendix.

5 Discussion

Our findings indicate that pragmatic, semantic, and syntactic ambiguities, affecting the control flow, have a strong impact on cognitive and behavioral aspects. They raise readers' cognitive load, lower their comprehension, and cause a significant amount of visual associations, implying increased cognitive integration effort [8]. Conversely, lexical ambiguities, affecting the labels, have a less pronounced impact. Effects could be observed in terms of lower comprehension and higher visual associations, but not clearly in terms of cognitive load. In particular, the mean number of fixations with duration $\geq 250ms$ (associated with mental processing, cf. Sect. 2.2) did not significantly differ between models with and without ambiguities. A possible interpretation is that ambiguities in the lexicon do not impact cognitive load as much as ambiguities in the control flow.

Having demonstrated the adverse effects of ambiguities, the question arises on how to better support model readers either by minimizing the presence of ambiguities in a process model or by providing better support in handling them.

For example, automated techniques for label refactoring have the potential to help to reduce lexical ambiguity in the first place (cf. [34]). However, we cannot assume that ambiguities can be completely removed, and acknowledge that certain ambiguities are intentional, e.g., to allow for flexible process interpretation and execution (cf. [15]). Nevertheless, the impact of ambiguities on task performance necessitates support that goes beyond the detection of modeling errors. This could translate to promoting a feedback loop in which modelers and model readers closely cooperate to identify ambiguous process elements. Another possibility is process model analysis tools that automatically detect ambiguities based on formal definitions such as those in Sect. 2.1 or in [15], possibly with the support of ontology annotations akin to [12]. Another possibility is that, with further development, the cognitive and behavioral measures captured with eye-tracking could be used to develop context adaptive systems that use pre-trained machine learning models (e.g., [6]) to detect when users are facing ambiguities and help them to disambiguate the model. This can, for instance, be done by showing additional artifacts such as textual annotations or guided simulations.

Our findings have implications for research, practice, and education. As for research, future empirical studies on human and cognitive aspects in process modeling should consider that ambiguity in process models may be a confounding factor. Therefore, it should be mitigated in the design of future experiments. As for practice, our results emphasize the need to strive for process models minimizing ambiguities and to provide disambiguation cues along with the models, e.g., enriching process models with additional artifacts (cf. [2]). As for education, our results call for increasing awareness in process modeling trainees about potential ambiguities in their models and about the effects of these ambiguities.

We acknowledge possible threats to the validity of our study. First, the presence of confounding factors that cannot be fully excluded threatens *internal validity*. However, this threat was mitigated through the design of a controlled experiment following a pre-defined research model (cf. Sect. 3.1), a careful and systematic preparation of the experiment material (cf. Sect. 3.1), a randomization of the presentation order of tasks to avoid learning and fatigue effects (cf. Sect. 3.2) and the use of a strict data collection protocol that was executed uniformly during all the data collection sessions (cf. Sect. 3.2). An additional threat to internal validity derives from asking the participants to state whether they had any issues in performing each task after completing it. Since each task involved a process model specifically designed to include an ambiguity hindering the task execution, it is possible that, nudged by the question, the participants could eventually anticipate the presence of issues in each new task. This anticipation constitutes a potential confounding factor as it may have an influence at the cognitive level. To mitigate this threat, we did not disclose the study goals to the participants, avoided mentioning that they would be facing issues while analyzing the process models, avoided making them aware of the concept of ambiguity in process models, and never mentioned ambiguity. We also kept the question intentionally generic to avoid hinting at the fact that participants were analyzing ambiguous process models. Additionally, having process models composed of one ambiguous and two non-ambiguous sub-processes allowed us to not always present the participants with models with issues. Finally, the randomized presentation of tasks (cf. Sect. 3.1) allowed us to uniformly spread across all tasks any increased cognitive load due to the potential anticipation of issues. Nevertheless, in future studies we could further mitigate this threat by including tasks involving models free from ambiguity and by phrasing questions avoiding hinting terms such as *issues*, to minimize the risk that participants may anticipate and look for issues. The inability to generalize the experiment results due to the participants' sample size or the used modeling language may threaten the *external validity* of our study. To mitigate these threats, we recruited 44 participants, which, to the best of our knowledge, places our study among the most extensive eye-tracking studies in process modeling. As for the use of BPMN, we argue that the tested ambiguities are not BPMN-specific but could be found in other imperative modeling languages such as workflow nets and EPC models.

6 Conclusion

Ambiguities in process models lead to multiple possible process interpretations. Using eye-tracking, we investigated how these ambiguities affect the model reader while conducting various model comprehension tasks. Our results demonstrate a significant impact on cognitive load, comprehension, and visual associations.

The demonstrated adverse effects of ambiguities suggest the need for adequate training and care in avoiding ambiguities in process models. Moreover, they motivate the need for the design of automated tools to aid model readers in detecting ambiguities and to possibly disambiguate them. In future work, we will further investigate the role of BPM expertise in detecting ambiguities (also in relation to diverse industrial backgrounds), strategies and behavioral patterns for handling ambiguities, and the specific impact of intentional ambiguities.

Acknowledgements. This work has received funding from the Swiss National Science Foundation under Grant No. IZSTZ0_208497. A. Abbad-Andaloussi is supported by the International Postdoctoral Fellowship Grant (No. 1031574) from the University of St.Gallen, Switzerland. H. A. López is supported by research grant "Center for Digital CompliancE (DICE)" (VIL57420) from VILLUM FONDEN. ChatGPT was used to polish sentences and correct spelling.

References

1. Empirical standard guidelines for experiments, `https://github.com/acmsigsoft/EmpiricalStandards/blob/master/docs/Experiments.md`
2. Abbad-Andaloussi, A., Burattin, A., Slaats, T., Kindler, E., Weber, B.: On the declarative paradigm in hybrid business process representations: A conceptual framework and a systematic literature study. Inf Syst **91**, 101505 (2020)
3. Abbad-Andaloussi, A., Burattin, A., Slaats, T., Kindler, E., Weber, B.: Complexity in declarative process models: Metrics and multi-modal assessment of cognitive load. Expert Systems with Applications **233**, 120924 (2023)

4. Abbad-Andaloussi, A., Burattin, A., Slaats, T., Petersen, A.C.M., Hildebrandt, T.T., Weber, B.: Exploring the understandability of a hybrid process design artifact based on dcr graphs. In: BPMDS 2019, Proceedings. pp. 69–84. Springer (2019)
5. Abbad-Andaloussi, A., Lübke, D., Weber, B.: Conducting eye-tracking studies on large and interactive process models using eyemind. SoftwareX **24**, 101564 (2023)
6. Abbad-Andaloussi, A., Sorg, T., Weber, B.: Estimating developers' cognitive load at a fine-grained level using eye-tracking measures. In: International Conference on Program Comprehension. pp. 111–121 (2022)
7. Amna, A.R., Poels, G.: Ambiguity in user stories: A systematic literature review. Information and Software Technology **145**, 106824 (2022)
8. Bera, P., Soffer, P., Parsons, J.: Using eye tracking to expose cognitive processes in understanding conceptual models. MIS quarterly **43**(4), 1105–1126 (2019)
9. Campbell, D.J.: Task complexity: A review and analysis. Academy of management review **13**(1), 40–52 (1988)
10. Corradini, F., Ferrari, A., Fornari, F., Gnesi, S., Polini, A., Re, B., Spagnolo, G.O.: A guidelines framework for understandable bpmn models. Data & Knowledge Engineering **113**, 129–154 (2018)
11. Dijkman, R.M., Dumas, M., Ouyang, C.: Semantics and analysis of business process models in bpmn. Information and Software technology **50**(12), 1281–1294 (2008)
12. Fan, S., Hua, Z., Storey, V.C., Zhao, J.L.: A process ontology based approach to easing semantic ambiguity in business process modeling. Data & Knowledge Engineering **102**, 57–77 (2016)
13. Figl, K.: Comprehension of procedural visual business process models. Business & Information Systems Engineering **59**(1), 41–67 (2017)
14. Figl, K., Mendling, J., Strembeck, M.: The influence of notational deficiencies on process model comprehension. J. Assoc. Inf. Syst. **14**(6), 1 (2013)
15. Franceschetti, M., Seiger, R., López, H.A., Burattin, A., García-Bañuelos, L., Weber, B.: A characterisation of ambiguity in bpm. In: International Conference on Conceptual Modeling. pp. 277–295. Springer (2023)
16. Glöckner, A., Herbold, A.K.: Information processing in decisions under risk: Evidence for compensatory strategies based on automatic processes. MPI collective goods preprint (2008/42) (2008)
17. Gulden, J., Burattin, A., Abbad-Andaloussi, A., Weber, B.: From analytical purposes to data visualizations: a decision process guided by a conceptual framework and eye tracking. Software and Systems Modeling **19**, 531–554 (2020)
18. Haisjackl, C., Pinggera, J., Soffer, P., Zugal, S., Lim, S.Y., Weber, B.: Identifying quality issues in bpmn models: an exploratory study. In: Int. Workshop on Business Process Modeling, Development and Support. pp. 217–230. Springer (2015)
19. Heggset, M., Krogstie, J., Wesenberg, H.: The influence of syntactic quality of enterprise process models on model comprehension. In: CAiSE Forum. pp. 89–96 (2015)
20. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., van de Weijer, J.: Eye Tracking: A comprehensive guide to methods and measures. OUP Oxford (2011)
21. Kindler, E.: Model-based software engineering and process-aware information systems. Transactions on Petri Nets and Other Models of Concurrency II: Special Issue on Concurrency in Process-Aware Information Systems pp. 27–45 (2009)
22. Krogstie, J.: Model-based development and evolution of information systems: A Quality Approach. Springer Science & Business Media (2012)

23. Leopold, H., Smirnov, S., Mendling, J.: Refactoring of process model activity labels. In: 15th Int. Conf. on Applications of Natural Language to Information Systems. Proceedings 15. pp. 268–276. Springer (2010)
24. Mendling, J., Recker, J., Reijers, H.A.: On the usage of labels and icons in business process modeling. Int. J. Inf. Syst. Model. Des. (IJISMD) **1**(2), 40–58 (2010)
25. Mendling, J., Reijers, H.A., van der Aalst, W.M.: Seven process modeling guidelines (7pmg). Information and software technology **52**(2), 127–136 (2010)
26. Mendling, J., Reijers, H.A., Recker, J.: Activity labeling in process modeling: Empirical insights and recommendations. Information Systems **35**(4), 467–482 (2010)
27. Mendling, J., Strembeck, M., Recker, J.: Factors of process model comprehension—findings from a series of experiments. Decision Support Systems **53**(1), 195–206 (2012)
28. OMG: Business Process Model and Notation (BPMN), Version 2.0.2. 2013, `https://www.omg.org/spec/BPMN/2.0`
29. Oppl, S.: Articulation of work process models for organizational alignment and informed information system design. Inf. Manag. **53**(5), 591–608 (2016)
30. Petre, M.: Why looking isn't always seeing: readership skills and graphical programming. Communications of the ACM **38**(6), 33–44 (1995)
31. Petrusel, R., Mendling, J.: Eye-tracking the factors of process model comprehension tasks. In: CAiSE 2013, Proceedings 25. pp. 224–239. Springer (2013)
32. Petrusel, R., Mendling, J., Reijers, H.A.: Task-specific visual cues for improving process model understanding. Inf. Softw. Technol. **79**, 63–78 (2016)
33. Petrusel, R., Mendling, J., Reijers, H.A.: How visual cognition influences process model comprehension. Decision Support Systems **96**, 1–16 (2017)
34. Pittke, F., Leopold, H., Mendling, J.: Automatic detection and resolution of lexical ambiguity in process models. IEEE Trans. Softw. Eng. **41**(6), 526–544 (2015)
35. Pustejovsky, J.: The generative lexicon. MIT press (1998)
36. Schreiber, C., Abbad-Andaloussi, A., Weber, B.: On the cognitive effects of abstraction and fragmentation in modularized process models. In: Business Process Management: 21st International Conference, BPM 2023 (2023)
37. Sweller, J.: Cognitive load theory. In: Psychology of learning and motivation, vol. 55, pp. 37–76. Elsevier (2011)
38. Veltman, J., Jansen, C.: The role of operator state assessment in adaptive automation. TNO Defence, Security and Safety Soesterbery (2005)
39. Wang, W., Chen, T., Indulska, M., Sadiq, S., Weber, B.: Business process and rule integration approaches—an empirical analysis of model understanding. Information Systems **104**, 101901 (2022)
40. Weber, B., Fischer, T., Riedl, R.: Brain and autonomic nervous system activity measurement in software engineering: A systematic literature review. Journal of Systems and Software p. 110946 (2021)
41. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin **1**(6), 80–83 (1945)
42. Winter, M., Neumann, H., Pryss, R., Probst, T., Reichert, M.: Defining gaze patterns for process model literacy–exploring visual routines in process models with diverse mappings. Expert Systems with Applications **213**, 119217 (2023)
43. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in software engineering. Springer Science & Business Media (2012)
44. Zugal, S.: Applying Cognitive Psychology for Improving the Creation, Understanding and Maintenance of Business Process Models. Ph.D. thesis, University of Innsbruck (2013)