Software and Systems Modeling manuscript No.
(will be inserted by the editor)

Exploring How Users Engage with Hybrid Process Artifacts Based on Declarative Process Models

A behavioral analysis based on eye-tracking and think-aloud

Amine Abbad Andaloussi · Francesca Zerbato · Andrea Burattin · Tijs Slaats · Thomas T. Hildebrandt · Barbara Weber

Received: date / Accepted: date

Abstract *Context:* Process design artifacts have been increasingly used to guide the modeling of business processes. To support users in designing and understanding process models, different process artifacts have been combined in several ways leading to the emergence of the so-called "hybrid process artifacts". While many hybrid artifacts have been proposed in the literature, little is known about how they can actually support users in practice.
Objective: To address this gap, this work investigates the way users engage with hybrid process artifacts during comprehension tasks. In particular, we focus on a hybrid representation of DCR graphs (DCR-HR) combining a process model, textual annotations and an interactive simulation.
Method: Following a qualitative approach, we conduct a multi-granular analysis exploiting process mining, eye-tracking techniques, and verbal data analysis to scrutinize the reading patterns and the strategies adopted by users when being confronted with DCR-HR.
Results: The findings of the coarse-grained analysis provide important insights about the behavior of domain experts and IT specialists and show how user's background and task type change the use of hybrid process artifacts. As for the fine-grained analysis, user's behavior was classified into goal-directed and exploratory and different strategies of using the interactive simulation were identified. In addition, a progressive switch from an exploratory behavior to a goal-directed behavior was observed. These insights pave the way for an improved development of hybrid process artifacts and delineate several directions for future work.

Keywords Process Models · Hybrid Process Artifacts · DCR graphs · Eye-Tracking · Think-aloud · Behavioral Analysis

1 Introduction

The design and development of Process-Aware Information Systems (PAIS) encompasses the creation of several process design artifacts (process artifacts, for short) aimed to support users in modeling, enacting and managing business processes. Such artifacts may include constructs, models, methods and instantiations [1], which are created to support users in solving a specific problem throughout the different phases of the business process life-cycle.

Over the years, process artifacts have been more and more integrated in the development of PAIS, leading to

---

Work supported by the Innovation Fund Denmark project *EcoKnow* (7050-00034A); the fourth author additionally by the Danish Council for Independent Research project *Hybrid Business Process Management Technologies* (DFF-6111-00337).

Amine Abbad Andaloussi
Technical University of Denmark, Denmark
E-mail: amab@dtu.dk

Francesca Zerbato
University of St. Gallen, Switzerland
E-mail: francesca.zerbato@unisg.ch

Andrea Burattin
Technical University of Denmark, Denmark
E-mail: andbur@dtu.dk

Tijs Slaats
University of Copenhagen, Denmark
E-mail: slaats@di.ku.dk

Thomas T. Hildebrandt
University of Copenhagen, Denmark
E-mail: hilde@di.ku.dk

Barbara Weber
University of St. Gallen, Switzerland
E-mail: barbara.weber@unisg.ch

hybrid solutions that loosely combine different artifacts with the aim to support the design and comprehension of process models [2,3], especially those lying under the umbrella of the declarative paradigm [4–9]. Such hybrid solutions combine (graphical) process models with textual process specifications [6, 8] and interactive simulations [4, 10].

In this work, we refer to *hybrid process artifacts* as representations combining two or more design artifacts (e.g., process models, textual annotations or interactive simulations) overlapping in the description of some business process aspects [11] and we specifically consider declarative process models [12].

In literature, there have been several hybrid process artifacts proposed to tackle the notorious limitations of declarative process models, particularly with regard to their understandability and maintainability [4,5,13,14]. Indeed, to support flexibility in process design, declarative process models capture constraints on the allowed activity flows. Hence, their interpretation requires a constant awareness of the states of all the constraints in the model throughout process execution, which can cause a high burden when being confronted with complex models [15]. This challenge also affects the maintainability of declarative process models. Herein, the support of a hybrid process artifact can ease the interpretation of the process model and enhance its maintainability.

While previous research has focused on proposing hybrid process artifacts to support users in designing and understanding process models [5,7,16], at this time, there is a lack of empirical research about how these hybrid representations are used. In particular, an in-depth understanding of how users engage with hybrid process artifacts is needed. Besides, an understanding of the benefits and challenges associated with the use of the different artifacts is required to contextualize the observed behaviors. Moreover, it is unclear whether users follow certain strategies when engaging with hybrid process artifacts.

To address this gap, the goal of this paper is to provide insights on how people engage with hybrid process artifacts. More specifically, this paper investigates how people use a specific hybrid artifact during comprehension tasks. In particular, we focus on the hybrid representation of DCR Graphs (DCR-HR, for short). This hybrid process artifact combines a declarative process model represented as a Dynamic Condition Response (DCR) graph [17], its textual specification and an interactive simulation. DCR Graphs are a well-known declarative process modeling language based on directed graphs whose nodes represent events and whose edges capture the relationship between them [17]. DCR Graphs benefit from the support of the DCR Graphs Portal [18], a research-based commercial tool supporting the design, enactment and analysis of DCR Graphs. The DCR Graphs Portal features a graphical web-based editor, a textual process specification that can be visually linked to parts of the process model and an interactive simulator that can be directly enabled from the editor and visualizes the process execution directly on the DCR Graph. Being widely adopted by industrial and governmental institutions in Denmark, the DCR Graphs Portal and, more in general, DCR Graphs are a valuable candidate for user behavior studies compared to solutions based on other declarative process modeling languages such as DECLARE [15] that have not been commercialized so far.

To gain understanding of how people engage with DCR-HR, we designed an exploratory study asking people to perform a set of comprehension tasks using DCR-HR. The goals of the study are to (i) observe the distribution of attention among the different artifacts considering different groups of stakeholders and different kinds of comprehension tasks, (ii) gather insights on the perceived benefits and challenges associated to each artifact, and (iii) identify common strategies describing how people approach comprehension tasks and use different artifacts over time.

For collecting data we rely on two different well-known approaches, namely eye-tracking [19] and retrospective think-aloud. Then, we favor a qualitative data analysis approach to explore user's behavior.

Eye-tracking has been applied to numerous fields in order to understand the complete user experience during the execution of different tasks [20], as it provides insights on the natural interaction of a user with a system [21]. In this paper, we collect eye-tracking data of users engaging with DCR-HR and analyze them qualitatively following two distinct approaches. A first coarse-grained analysis exploits process mining techniques [22] and attention maps [19] to investigate how the three different artifacts of DCR-HR are used, specifically focusing on how groups of stakeholders with different backgrounds relate to DCR-HR while performing different kinds of tasks. As our focus is on attention distribution, during this analysis phase we are interested in gaining insights into *how much* each artifact is used individually and in combination with others. Then, we conduct a more fine-grained analysis considering the different elements of the DCR Graph individually and exploring *how* artifacts are used *over time*. During this analysis, we look deeper into temporal patterns and observe common strategies describing how people engage with DCR-HR during the comprehension task. In this analysis, we rely on scarf-plot visualizations [23] and, then, build on the well-known dichotomy of goal-directed and exploratory search behaviors [24, 25] to categorize the identified strategies.

Retrospective think-aloud is a research method used to verbalize users' thoughts after the execution of a certain task [19]. In this paper, we use retrospective think-aloud to extract the subjective insights of the participants who took part in the exploratory study, focusing on perceived benefits and challenges associated with the artifacts of DCR-HR. This allows us to obtain explicit user feedback, which is needed for having a complete picture of *why* users behaved in a certain way and for enhancing the interpretation of eye-tracking data during the described analyses.

In general, using behavioral data and think-aloud can inform us on the use of hybrid process artifacts (through implicit and explicit feedback [26]). Overall, the results of the exploratory study suggest that different groups of stakeholders tend to use different artifacts of DCR-HR and that usage changes based on the type of task being executed. Indeed, the users' background seems to affect perceived benefits and challenges, thus influencing the way different artifacts are used for achieving a specific purpose. When examining the use of artifacts over time, we found that users follow different strategies to interact with the artifacts of DCR-HR and observed different ways in which the interactive simulation was used. In addition, we noticed that people tend to switch from an exploratory behavior to goal-directed behavior progressively. The outcomes of this study confirm that the use of different process artifacts enhances the experience of users with different backgrounds, also based on the kind of task being executed. Besides, our findings pave the path for future research in the direction of improving the design of hybrid process artifacts, for example by considering explicit user preferences and implicit feedback to add or eliminate certain artifact features.

This paper extends original work initially presented in [10] by providing a broader and more complete overview of how users engage with DCR-HR. In particular, we introduce a novel fine-grained analysis which considers DCR Graphs and centers around the temporal dimension of the eye-tracking data. By observing how the users' behavior unfolds over time, we are able to identify interesting strategies describing how users engage with DCR-HR.

The remainder of this paper is organized as follows. Section 2 provides the reader with useful background concepts. Section 3 introduces related work. Section 4 presents the research method followed to design the exploratory study. Section 5 reports the results of the analysis. Section 6 discusses the main findings and highlights the interesting outcomes of this research, as well as its limitations. Finally, Section 7 concludes the paper and delineates the directions for future work.

2 Background

This section presents the main notions employed throughout this paper. We start with a general introduction to hybrid process artifacts (Section 2.1), followed by a description of DCR-HR, (Section 2.2). Finally, we introduce eye-tracking as one of the core methodologies behind our study (Section 2.3).

2.1 Hybrid Process Artifacts

Hybrid process artifacts combine two or more process artifacts (e.g., process models, textual annotations or interactive simulations) overlapping in the description of some business process aspects [11]. Hybrid process artifacts have been proposed in the literature to address several challenges within the areas of process modeling (for a systematic literature review see [11]), in particular to address open challenges in the use of declarative languages [27,28] caused by their limited understandability [29] and maintainability [30]. The limited capacity of humans when dealing with constraints is among the key challenges in that respect. Indeed, a full understanding of a declarative process model requires being aware of the states of all the constraints in the model throughout the whole process execution [28]. This requirement gets more complicated when considering the implicit constraints (also called "hidden dependencies [5]) between the model activities and all the possible ways in which they could interact. As the capacity of the human working memory is usually limited to 7 ± 2 items [31], interpreting declarative process models with too many constraints, without the support of others process artifacts (e.g., interactive simulations or textual annotations), becomes a challenging task. The limitations of declarative languages go beyond their understandability, as their maintainability is also quickly hampered by hidden dependencies. Considering the entanglement of hidden dependencies and the abundance of ways in which they could interact, it becomes challenging to infer the set of constraints affected by a change of the process specifications and to ensure the consistency of the model after altering some of its constraints. Hence, without the support of additional artifacts, the maintainability of a declarative process model is prone to misalignment and non-compliance with the process specification.

2.2 DCR-HR: A Hybrid Representation of DCR Graphs

DCR-HR is a hybrid process artifact combining (i) a DCR Graph [17], with (ii) textual process specifications and (iii) a simulation allowing to evaluate the behavior of the process model. DCR Graphs and DCR-HR have been developed through a close collaboration between academia and industry [32,33], combining research into formal methods and declarative notations with the development of a commercial modelling tool[1] and its application to real-world cases [34].

The inclusion of textual process specifications in the presentation was driven by the ongoing EcoKnow research project[2], where the DCR technology is being applied to support the effective digitization of citizen processes. The use of DCR-HR, combining textual legal paragraphs with the graphical DCR notation and guided simulation, is used to empower knowledge workers in Danish municipalities by enabling them to make sense of digitized models of the law.

The version of DCR-HR considered within this paper is based on this application within the EcoKnow project and specifically refers to a process derived from section §45 of the Danish "Consolidation Act on Social Services" [3]. Its layout is depicted in Fig. 1. Here, we have on the left a DCR Graph modelling the aforementioned process and consisting of: (i) boxes denoting the activities of the process, e.g. *offer 15 hours of assistance* and *designate a person*, and (ii) arrows between the boxes denoting the constraint on the process, e.g the yellow arrow with a dot on the end between the former two activities. The activities can be assigned a role, placed in the bar above the box, e.g. *Receiver* for *designate a person*. They can also be nested inside each other, indicating a form of hierarchy, e.g. all activities are a part of *paragraph 45*. When constraints are drawn between such nestings they apply to all child activities. As is the norm for declarative notations, unconstrained activities can be executed freely, i.e., at any time and any number of times. DCR Graphs include five types of constraints: (i) the *condition*, drawn as a yellow arrow with a dot on the end, denotes that before one activity can be executed, another needs to have been done at least once in the past; (ii) the *response*, drawn as a blue arrow with a dot at the start, denotes that after one activity is executed, some other activity becomes required and needs to be done before the process can be finalized; (iii) the *exclusion*, drawn as a red arrow with a percentage sign at the end, denotes that when one activity is executed another activity is removed of the process; (iv) the *inclusion*, drawn as a green arrow with a plus sign at the end, denotes that when one activity is executed another activity is added back into the process; finally (v) the *milestone*, drawn as a purple arrow with a diamond at the end, denotes that while one activity is required to be done another activity is blocked from executing.

At the bottom of Fig. 1 we have the law text, taken directly from the relevant laws that govern the process. Each fragment of the law text is linked to the currently selected activity: by selecting *paragraph 45* users will be able to visualize the entire law text for the current process, while by selecting *designate a person* they will see only the part of the law that is relevant for this activity.

Finally, at the right side of Fig. 1, we have the interactive simulation. This consists of (i) a clickable list of currently executable tasks, (ii) a textual log of what has already been simulated, (iii) a swim-lane representation of this log. Each time an activity is executed for simulation, it will be added to the textual log and swim-lane, and the list of currently executable tasks will update.

The combination of these different artifacts is meant to allow users to form their own strategies when reading the model, based on their personal background and the type of comprehension tasks that they are trying to solve. For example, when asked a question about the law some users may use the graphical process model to find the answers, while others may focus on the text. For highly operational questions, some users may prefer to use the guided simulation, as it allows them to quickly try out different execution scenarios without having to memorize all the possible execution traces.

2.3 Eye-tracking

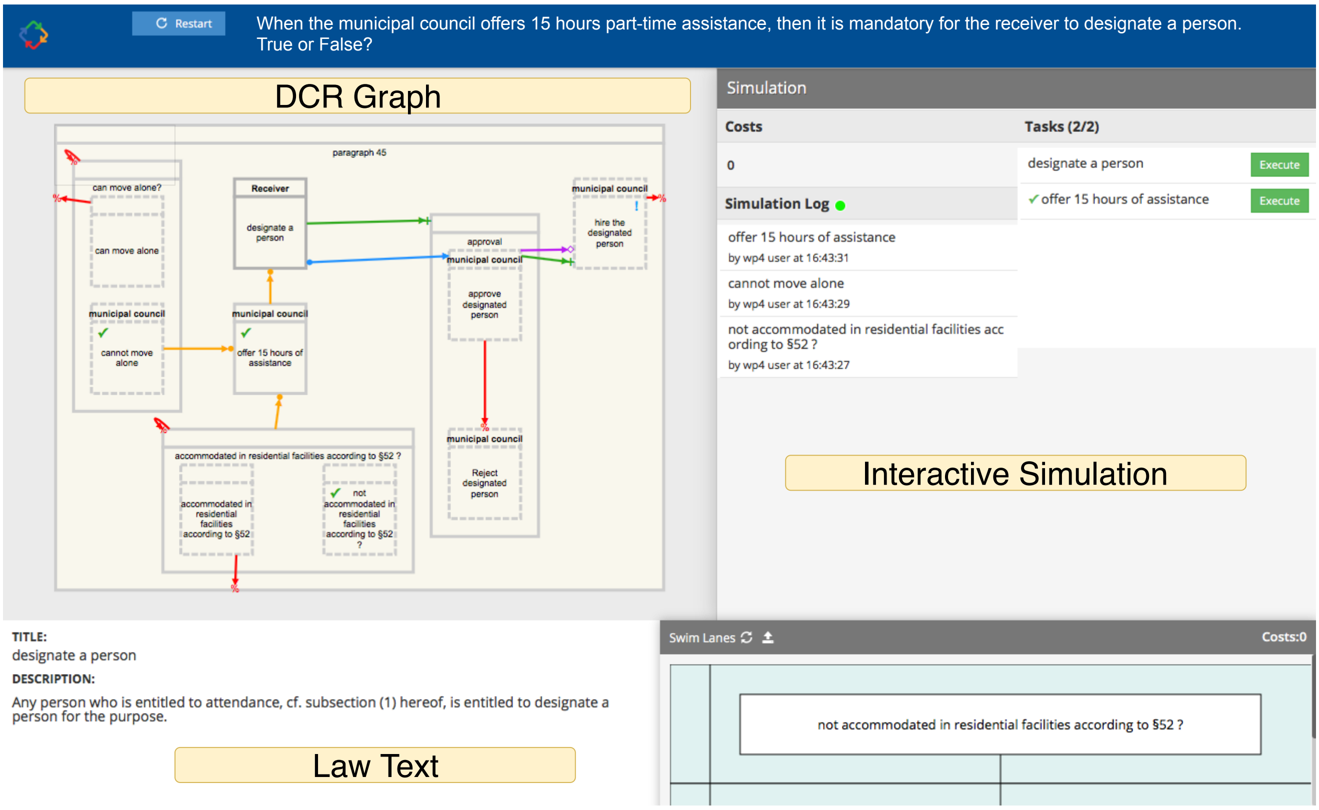
Eye-tracking is a widely adopted methodology allowing researchers to track humans' gaze interactions with external stimuli [20]. Eye movements are recorded by eye-tracking devices as gaze data, which in turn are used to derive a set of oculomotor events such as *fixations* and *saccades*. The detection of these events is associated with a set of properties such as duration, amplitude, and velocity [19]. A fixation refers to the time span when the eye remains still at a specific position of the stimulus [19]. An example of fixation is the time the eye stops at a word while reading a sentence. A saccade refers to the rapid eye movement occurring between fixations [19]. When reading a sentence, saccades occur when the reader moves from one word to another.

---

[1] `http://www.dcrgraphs.net`

[2] `https://ecoknow.org`

[3] `http://english.sm.dk/media/14900/consolidation-act-on-social-services.pdf`



When the municipal council offers 15 hours part-time assistance, then it is mandatory for the receiver to designate a person.
True or False?

offer 15 hours of assistance
by wp4 user at 16:43:31

cannot move alone
by wp4 user at 16:43:29

not accommodated in residential facilities according to §52 ?
by wp4 user at 16:43:27

TITLE:
designate a person

DESCRIPTION:
Any person who is entitled to attendance, cf. subsection (1) hereof, is entitled to designate a person for the purpose.

not accommodated in residential facilities according to §52 ?

Fig. 1 A view showing the DCR-HR layout. The hybrid process artifact comprises a DCR Graph, law text and an interactive simulation.

The availability of fixations and saccades allows for a wide range of statistical and visual analyses of human behavior. While statistical analyses (i.e., descriptive statistics, inferential statistics, and statistical modeling) provide the basis for hypothesis testing, their use requires aggregating data to a level where the application of statistical methods is possible. However, this comes at the cost of providing rather coarse and limited insights about the human visual behavior [35]. In addition, the interpretation of the metrics inferred from the aggregated data is tightly coupled with the context and the application where these metrics have evolved. As a result, the mapping between the numerical data and the real perceptual and cognitive features of the participants remains uncertain.

Alternatively, visual analyses follow an exploratory approach to investigate the way people approach a stimulus. Graphical plots allow visualizing the spatial, temporal and spatio-temporal features of eye-tracking data [35]. Many of these representations are based on the notion of *scan-path*, i.e., the path of oculomotor events in space recorded during a certain period of time [19].

However, with the increasing complexity of the stimulus and the high number of fixations recorded throughout several eye-tracking trials, typical visualizations become cluttered and hard to understand. To overcome this issue, the stimulus can be segmented into *Area(s) of Interest* (AOI), which define the regions of the stimulus that are susceptible to provide meaningful insights to researchers. These regions are usually used to investigate the focus on specific parts of the stimulus [19]. Eye-tracking data can be grouped by areas of interest and represented as *dwells*. A dwell refers to one visit in an AOI from entry to exits points [19]. A common metric associated to AOIs is the dwell time, i.e., the time gazed at a certain AOI, from entry to exit [36], including all fixations and saccades landing within the AOI. The total dwell time is the sum of all dwell times to a specific AOI over a trial, including revisits on the same AOI. Representing eye-tracking data as dwells allows describing the visual behavior in a more concise way, based on the visited AOIs.

The eye-tracking literature proposes different techniques to visualize the way people engage with visual stimuli [23]. These visualizations provide insights about common strategies and patterns followed by the humans when interacting with the stimulus. In the context of this study, we deploy two visualizations: attention maps and scarf-plots. Attention maps [37] are aggregation of fixations over time and/or over participants [23] and are often used as an alternative for transitions matrices to capture transitions between AOIs. Unlike transition matrices that emphasize only the spatial feature of AOIs [38], attention maps consider also the ordering dependencies between AOIs. To obtain these maps, fixations are grouped by dwells, then direct relationships between dwells are identified and used to construct a dependency graph illustrating reading patterns. Attentions maps provide important insights about the dwell time and the transitions between AOIs. However, the generated dependency graphs are unable to describe the evolution of the reading patterns over time. Instead, scarf-plots [19] provide a timeline representation allowing to observe the changes in the visited AOIs over time, which in turn could provide pertinent insights about the users' visual search behavior.

The literature discerns two types of search behaviors, namely *goal-directed search* (also referred to as "task-oriented processing" in [39]) and *exploratory search* [20, 24]. Goal-directed search occurs when users follow a certain search routine to gather general or specific information to solve a task. Oppositely, exploratory search occurs when users lack the experience to search efficiently and rather screen the environment without having a clear search plan in mind [20]. According to a recent study about consumers' attention processes [25], the total number of different products that are fixated by a user and the average time that the user spends fixating detailed product information can be used to distinguish between goal-directed and exploratory search. The comprehension of process models comprises an active search for relevant information [40], which would be interesting to scrutinize in the light of the search behaviors reported in the literature.

3 Related Work

This section presents the works related to this paper. Section 3.1 introduces existing research on hybrid process artifacts, Section 3.2 describes the eye-tracking studies in the context of process modeling, and Section 3.3 summarizes behavioral analyses conducted in the field process modeling. Finally, Section 3.4 highlights the novelty of our work compared to previous research.

3.1 Related Studies on Hybrid Process Artifacts

The literature comprises a wide range of studies evaluating the understandability (e.g., [41]), maintainability (e.g., [42]) and modeling (e.g., [43]) of business processes. However, with regards to hybrid process artifacts, only a handful of empirical studies exist (for a systematic literature review see [11]). In [4, 5], Zugal et al. introduce the Test Driven Modeling (TDM) approach. This hybrid approach combines a declarative process model (modeled in Declare [28]) with test cases to support users when interpreting declarative models. The hybrid process artifact proposed by the authors was evaluated in two empirical studies [44]. The results show that the use of test cases as a hybrid process artifact provides an additional communication channel and supports the maintainability of declarative process models by lowering the participants' cognitive load and increasing the perceived model quality.

De Smedt et al. [6, 7] propose a hybrid process artifact combining a declarative process model with textual annotations describing the hidden dependencies in the model. In this work, the authors address the issue of hidden dependencies and propose a methodology allowing to infer them and make them explicit for the users. Following a quantitative approach, the authors evaluate the understandability of the proposed hybrid process artifact. The results support the hypothesis that the proposed hybrid representation contributes to better comprehension accuracy and reduced response time and cognitive load when compared to conventional (i,.e., non-hybrid) declarative representations.

Along the same lines, Lopez et al. [8] propose the process highlighter as a means to interlink the constraints in the process specifications with the constructs of the process model. The tool is embedded in a process artifact to support modelers during process modeling. The hybrid process artifact is evaluated by Andaloussi et al. [9]. Following a qualitative approach, the authors investigate the way users engage in a modeling task with the support of the process highlighter. The results suggest evident support during process modeling and highlight several benefits resulting from the explicit mapping between textual process specifications and modeling constructs, which in turn could improve the quality of designed process models.

3.2 Eye-Tracking in Process Modeling Research

Eye-tracking is used in different fields to study humans' behavior and cognition [19, 45]. In the field of process modeling, eye-tracking is used to study the understandability of process models and to predict users' performance in solving comprehension tasks. In particular, Petrusel and Mendling [46] investigate the impact of focusing on relevant regions on the comprehension of BPMN [47] process models. The authors formalize the notion of relevant regions, derive new metrics based on the notions of precision, recall, and their harmonic mean, and propose a statistical model allowing to predict the answer accuracy of participants based on such relevant regions.

Similarly, Bera et al. [48] study how attention on relevant parts of the process models influences users' performance, and scrutinize the visual association of modeling constructs during a comprehension task. Following a set of quantitative and qualitative studies supported by eye-tracking and concurrent think-aloud, these two research aspects are investigated considering different notations: BPMN [47], EPC [49] (Event-driven Process Chains) language, and EPC-H (i.e., a variant of EPC where roles are highlighted in different colors). The quantitative analysis shows that both pools and lanes in BPMN [47] and the highlighted roles in EPC-H help participants to focus on the relevant parts of the model, whereas EPC without role-highlighting fails to support that. As for the qualitative analysis, the results allow the authors to link the visual associations between different parts of the model (seen with the support of eye-tracking) and the underlying cognitive integration processes (articulated through concurrent think-aloud).

Empirical evaluations of process models using eye-tracking also cover models integrating business rules. Driven by the principle of separation of concerns and the need to integrate business rules in process models in an understandable manner, Wang et al. [50] investigate the effect of linked rules (i.e., a specific class of business rules) [51] on the understandability of process models following a quantitative approach. The outcome of the statistical analysis demonstrates that the integration of linked rules with process models is associated with decreased response time and reduced cognitive load, while a positive impact on the comprehension of the model remains partially supported.

The experience with the use of eye-tracking in process modeling is investigated by Zimoch et al. [52]. In this work, the authors report the lessons learned from a study investigating the comprehension of process models in different imperative notations. The findings are summarized into a set of recommendations meant to improve the design of eye-tracking experiments. Overall, these recommendations address the familiarity of participants with the process scenario and their prior expertise with the modeling language, which could potentially bias the validity of the results. In addition, the authors encourage combining eye-tracking measures with other cognitive biosensor-based measures and suggest that the results should be interpreted in light of the existing theories in cognitive psychology.

3.3 Behavior Analysis of Humans Engaging with Process Models

While several empirical studies investigate the comprehension of process models, only a few of them analyze the behavior of participants when interacting with these models. Haisjackl et al. [53] investigate the reading of declarative process models (in Declare [15]) following a qualitative approach supported by concurrent think-aloud. The analysis of the verbal protocols reveals important insights about the way the participants engage with the provided process models. In particular, the results show that the participants read Declare models in a sequential way, usually refer to the left corner activity to identify the entry-point of the model, follow a top-down strategy to read hierarchical processes, and ignore hidden dependencies when describing the model.

In another context, Haisjackl et al. [54] investigate the way imperative process models (expressed as BPMN) are inspected for quality checks following a qualitative approach supported by concurrent think-aloud. The analysis of the verbal data allows the identification of different strategies used for inspecting process models and sheds light on the order in which quality checks are generally performed.

Regarding the modeling of business processes, the PhD work of Pinggera [55] provides important insights about the behavioral patterns of modelers during process modeling. Following a qualitative approach supported by concurrent think-aloud and user interactions, the author identifies the order in which different modeling phases occur (e.g., problem understanding, method finding, modeling, reconciliation, validation) and proposes a catalog of modeling patterns. In a follow-up study, the author uses a quantitative approach to derive a statistical model supporting the identification of different modeling styles [56].

3.4 Novelty Compared to Previous Research

This paper differs from earlier studies in several aspects. As opposed to [6, 7, 44] our work takes a different approach to look at the comprehension of hybrid process artifacts by analyzing the participants' behavior when being confronted with hybrid process artifacts, which in turn allows perceiving the usability of hybrid process artifacts from a different perspective. With regards to [46], our interest does not lie in the definition or discovery of quantitative metrics for evaluating comprehension performance, but we rather focus on identifying general strategies that are descriptive of users' behavior in the context of process model comprehension. Moreover, while [46] focuses on quantitative findings, we base our analysis solely on qualitative data, complementing the findings derived from eye-tracking analysis with think-aloud data. Compared to [53,54], we consider eye-tracking data collection and analysis, thus adding a novel dimension to describe users' behavior. Last but not least, to the best of our knowledge, this is the first paper providing insights on how people read and interact with DCR Graphs alone and with DCR-HR.

4 Research Method

This section describes the steps followed to design the exploratory study and to analyze the collected data. In Section 4.1, we provide an overview about the design and the analysis of our study. Afterwards, in Section 4.2 we highlight the key aspects of the study design and execution, while in Section 4.3 we focus on the different approaches used for data analysis.

4.1 Overview

Our research aims at exploring how domain experts and IT specialists engage with hybrid process representations, in particular DCR-HR. To this end, we designed an exploratory study supported by eye-tracking and think-aloud and recruited municipal employees (serving as proxies for domain experts) and academics (serving as proxies for IT specialists). After the data collection, we conducted a multi-granular qualitative analysis to gain insights about the users' behavior when interacting with the different DCR-HR artifacts (i.e., DCR graph, law text, simulation).

Table 1 presents the goal of our study following the Goal, Question, Metric (GQM) template [57]. To address this goal we formulated three different research questions and used several eye-tracking measures and qualitative codes as indicators to understand the behavior of participants. Overall, we looked into the way participants with different background engage with DCR-HR when solving different tasks, assessed the benefits and challenges associated with each of the DCR-HR artifacts and explored the different search strategies used by the participants when interacting with DCR-HR. In the following (Sections 4.2 and 4.3), we describe the design and execution of our study and explain the different analysis approaches pursued to infer and use the measures and indicators presented in Table 1.

4.2 Exploratory Study Design and Execution

To gain insights into the use of hybrid process artifacts (specifically DCR-HR) during comprehension tasks, we design an exploratory study that collects and analyzes eye-tracking and retrospective think-aloud data.

Exploratory study design. As previously introduced, DCR-HR encompasses three different process artifacts, namely a DCR Graph [17], a textual process specification based on excerpts of the law and an interactive simulation (cf. Section 2.2). Such artifacts are meant to support stakeholders with different backgrounds when performing different types of tasks. Indeed, while domain experts are typically knowledgeable with the law, IT specialists are rather familiar with process models. Since little is known about how people engage with hybrid process artifacts, analyzing user behavior would provide important insights about the way the different DCR-HR artifacts are conjoined and used during a comprehension task. This, in turn, could help developers to improve the current tool based on personal user preferences. Moreover, literature [2,3,44] suggests that hybrid process artifacts have the advantage to provide different perspectives on the process which might be purposeful to a different extent depending on the specific task to be addressed. Thus, it is important to scrutinize whether they are used in the same way when performing different tasks. Such input is expected to further improve tool-support based on the characteristics of the running task and to provide ad-hoc recommendations to support each user based on his or her background. Based on these considerations we formulate the first research question as follows: **RQ1 - How do users engage with the different DCR-HR artifacts?**

In addition, to gain a deeper understanding of why users interact with DCR-HR in a certain way, we examine the perceived benefits and challenges associated with their use, keeping in mind that each artifact conveys information that can be used independently or combined with the one carried by other artifacts to pursue a certain goal. By collecting the subjective insights of the participants, we are also able to support the interpretation of the results related to RQ1. Hence, the second research question can be stated as follows: **RQ2 - What are the benefits and challenges associated with each one of the artifacts of DCR-HR?**

Finally, we are interested in exploring how the use of different artifacts evolves over to time. The temporal sequencing of eye movements allows us to identify common strategies followed to approach comprehen-

Goal	**Purpose** **Issue** **Object** **Viewpoint**	Explore the way users engage with DCR-HR from the domain experts and IT specialists viewpoints
Question **Metrics**	RQ1	How do users engage with the different DCR-HR artifacts? Mean fixation duration and mean transition frequency
Question **Indicators**	RQ2	What are the benefits and challenges associated with each one of the artifacts of DCR-HR? Qualitative codes emerging from the verbal utterances of participants
Question **Metrics**	RQ3	What strategies are followed when engaging with the different DCR-HR artifacts? Durations of fixations on relevant AOIs and total number of fixated AOIs

Table 1 GQM Model describing our goal, research questions and used metrics and indicators.

sion tasks. Accordingly, we formulate the third research question as follows: **RQ3 - What strategies are followed when engaging with the different DCR-HR artifacts?**

To explore the previously introduced research questions, we begun with collecting the gaze data of participants interacting with DCR-HR during process model comprehension tasks and then collected their subjective insights during a retrospective think-aloud session.

The context of this study is the digitalization of the law and, in particular, we focus on a process derived from paragraph §45 of the Danish "Consolidation Act on Social Services" (cf. Section 2.2).

Participants were 5 municipal employees from Syddjurs Municipality in Denmark and 10 academics studying or working at the Technical University of Denmark or at the IT University of Copenhagen. The municipal employees (who serve as proxies for domain experts) had proficiency in reading legal documents, but had (with one exception) no prior knowledge in process modeling or IT development. Academics (who reflect the common profiles of IT specialists) had background in process modeling and IT development, but lacked experience in reading and interpreting legal texts. Although, all academics have worked with process models in the past, not all of them were familiar with the DCR notation.

The data collection phase was organized into six comprehension tasks presented in the form of questions that participants had to answer using the different artifacts of DCR-HR. Following the input of experts in the field of process modeling, the comprehension tasks were designed to be easily grouped into three categories reflecting typical situations that a user may face when dealing with a process model.

The first group of *constraint tasks* comprises questions about the relationships between pairs of activities represented in the process model. These questions reflect a typical process comprehension task as a user is expected to have a clear understanding of which are the activities and the constraints that are relevant to the question to provide an answer. The second group of *decision tasks* comprises questions prompting the user to decide among multiple options. In this regard, a user is expected to identify the contextual information required to guide his or her decision-making process to achieve the desired outcome. Such contextual information is often not included in the process model and, thus, a user is expected to make a decision by relying on the process specification, i.e., the law text (cf. Section 2.2). Finally, *scenario tasks* concern the execution of partial process instances and comprise questions asking a user to determine whether or not a certain behavior is admissible based on a given case history. Scenario tasks are typical of process model testing and validation. Indeed, to validate a process based on positive and negative test scenarios, it is necessary to keep all execution traces in mind in order to check whether a certain behavior is admissible or not.

The comprehension tasks were designed and displayed on the DCR Graphs Portal [18] either in English or in Danish depending on the individual preference of the participants. Visually, the content of the web-based user interface was organized into the four main areas outlined in Fig. 1 (cf. Section 2.2).

The complete material designed for this study can be found online[4].

Executing the exploratory study. The execution of the exploratory study was organized into a PREPARATION PHASE and a DATA COLLECTION PHASE, which were carried out individually for each participant and whose details are captured by the BPMN [47] process of Fig. 2.

At the beginning of the study, each participant went through a PREPARATION PHASE consisting of the following four steps (cf. the first four tasks of the process of Fig. 2) a physical assessment questionnaire, a background data questionnaire, an introduction to DCR-HR

---

[4] `http://andaloussi.org/sosym2019/design/`

and to the DCR Graphs Portal, and a final calibration step.

The **Physical assessment questionnaire** was carried out to check the physical ability of a user to participate in a video-based eye-tracking experiment. Indeed, people wearing glasses or contact lenses, or having reduced vision may compromise the acquisition of high-quality eye-tracking data [36]. Overall, seven of the participants claimed to wear glasses or contact lenses and only one person (not among those wearing glasses) reported having some troubles seeing things located far away. However, since no-one reported any major vision issue, we decided not exclude anyone from the study at this point [36]. The physical assessment questionnaire and data are available in our online appendix.[5]

Afterwards, we administered a **Background and expertise questionnaire** asking each participant some questions about (i) his or her ability to read and create DCR Graphs and to use the interactive simulation, (ii) his or her familiarity with the process used in the comprehension tasks. With regards to (i) the collected data show that academics were indeed more familiar with reading, creating and simulating DCR Graphs than municipal employees (averages on a scale of 5: 3.9 and 1.8 familiarity with reading DCR Graphs respectively, 2.3 and 1.2 familiarity with creating DCR Graphs respectively, 3.6 and 1.2 familiarity with simulating DCR Graphs respectively). Concerning the familiarity with the considered process (ii) only one participant claimed to be already familiar with the considered process. The background and expertise questionnaire and data are available in our online appendix[6].

Then, we provided a 30-minute introduction to DCR-HR, going in detail through the semantics of the different DCR Graph relations [17], and presented the layout and the main features of the DCR Graphs Portal (cf. task **Introduction to DCR-HR and portal** in Fig. 2).

As a last preparation step, we conducted a hardware calibration procedure to ensure good data quality (cf. task **Calibration and testing** in Fig. 2). In detail, we used a 9-points calibration and tested its accuracy with all participants prior to starting with data collection.

The first comprehension task (i.e., **Familiarization task** in Fig. 2) was designed to allow participants get acquainted with the DCR Graphs Portal. Then, the remaining six comprehension tasks were displayed sequentially and the user could proceed in a self-paced way (participants took an average of 15 minutes to execute all the comprehension tasks). During the execution of the **COMPREHENSION TASKS**, we recorded eye gaze data using the Tobii Pro X3–120 eye-tracker[7], which was placed in front of the monitor that presented the comprehension tasks.

At the end, we conducted a **Retrospective think-aloud** session [19] to collect insights about the use of DCR-HR. During this phase, we recorded all the subjective insights provided verbally by the participants about their experience with DCR-HR. The session was organized as a guided interview structured as a set of open-ended questions meant to help the participants to articulate the adopted problem-solving strategies and identify the encountered challenges. The following two questions provide a good summary of what was asked to the participant during this last step of the data collection phase: (i) How and for what purpose did you use the following artifacts: model, law text, simulator? and (ii) Thinking about the overall experiment, was there anything challenging? The goal of the think-aloud was to encourage participants to elaborate on specific aspects that were deemed important for the exploratory study. Particularly, with (i) we tried to investigate the perceived benefits associated with each DCR-HR artifacts and to understand how the artifacts have been used and combined to perform different tasks. Similarly, with (ii) we invited participants to reflect on the challenges which could potentially hinder the use of the different DCR-HR artifacts. Both questions gave rise to important insights, which were coded and organized following the principles of grounded theory [58] to serve as a reference for the analysis introduced in the following section.

4.3 Analysis Approach

To answer research questions RQ1–RQ3, we followed different approaches to analyze the eye-tracking and think-aloud data collected during the exploratory study.

Eye-tracking data analysis. Eye-tracking is a widely used method to reveal patterns of visual behavior of humans [20] and provides insights into several information processing tasks [59].

Fig. 3 depicts the general ideas behind the eye-tracking data analysis followed in this paper.

Usually, the raw gaze data recorded by eye-tracking devices are aggregated into fixations and saccades, i.e., the two major categories of oculomotor events used in the analysis of eye movements [19]. In this study, we

---

[5] See `http://andaloussi.org/sosym2019/design/Physical_assessment_questionnaire.xlsx`

[6] See `http://andaloussi.org/sosym2019/design/Background_expertise_questionnaire.xlsx`

[7] See `https://www.tobiipro.com/product-listing/tobii-pro-x3-120/`

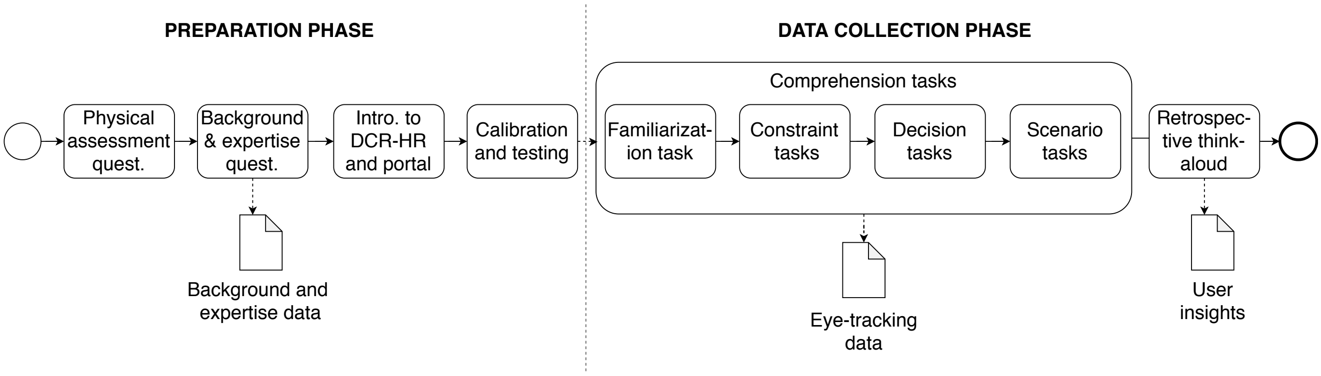


Fig. 2 BPMN process summarizing the PREPARATION and DATA COLLECTION phases of the conducted exploratory study. Collected data are captured by BPMN data objects.

derived fixation data from gaze points [19] using the I-VT Algorithm [60] implemented in the Tobii Pro Studio 3.4.8 software (cf. step ① in Fig. 3).

Then, prior to diving into the analysis, we looked into factors potentially affecting the quality of the collected data [36]. To this end, we replayed the eye gaze recordings to qualitatively assess the accuracy of the collected eye-tracking data for each individual. We also relied on the data quality measures provided by the eye-tracking software, e.g., the proportion of time spent on fixations compared to the total time spent executing the task. As a result of this data quality assessment step, we excluded two participants from the analysis. One participant, whom we labeled P04 was removed because of missing data: indeed, the average fixation time per comprehension task amounts to 02.53 seconds, which is much shorter than the 69.06 seconds spent by the other participants on average. This participant was the one claiming to have prior knowledge with the process, which may explain the reduced amount of fixations. Another participant, whom we labeled P10, reported having issues with peripheral view forcing him to rotate his head to read. Indeed, by replaying the gaze recordings, it is clear that the data collected for P10 are not accurate. Although the calibration was successful for both P04 and P10, it is likely that the mapping between their gazes and the corresponding coordinates on the stimulus failed at some point during the data collection.

Then, since our research questions consider different DCR-HR artifacts, which are located in a specific position of the screen, we labeled fixation data based on the targeted Area Of Interest (AOI). In this way, we were able to analyze the amount of time that each participant spent fixating a particular artifact of DCR-HR.

When defining areas of interest, we considered two levels of granularity. First, to answer **RQ1** at a coarse-grained level, we considered 3 areas of interest, each one referring to a distinct DCR-HR artifact, i.e., the DCR Graph, the law text and the simulation (cf. Fig. 1 and step (2a) in Fig. 3). Then, to enable a more fine-grained data analysis and look into users' behavior (cf. **RQ3**), we divided the AOI framing the DCR Graph into 22 smaller AOIs, each one referring to a distinct model element (i.e., activity or relation of the graph). We also defined a novel AOI including the question title, obtaining 25 AOIs overall as shown in Fig. 4 (cf. step (2b) in Fig. 3).

After having defined the AOIs, we exported from the eye-tracking software time-stamped data sets of fixations including an identifier for each participant, the considered comprehension question, the duration of the fixation and the AOI hit, which defines whether a fixation falls within a certain AOI or not [23] (cf. steps (3a) and (3b) in Fig. 3).

To analyze the obtained fixation data and explore research questions **RQ1** and **RQ3**, we followed two distinct approaches, based on the granularity used in the definition of the AOIs and on the analysis goal.

For the first analysis, we used the three coarser AOIs, exploited process mining [22] and AOI-based attention maps [19] to explore the relationships between different AOIs. This approach was followed to investigate **RQ1**.

As a first step, we transformed fixation data into an XES event log [61] by merging contiguous fixations referring to the same AOI (cf. step (4a) in Fig. 3). Then, after identifying the directly-follow relationships in the log [62], we generated a descriptive process model (referred to as "attention map" in the context of this work) illustrating how the attention of participants was distributed among the three artifacts of DCR-HR (cf. step (5a) in Fig. 3).

An example of such attention maps is depicted in Fig. 5. In the proposed graphical representation of at-

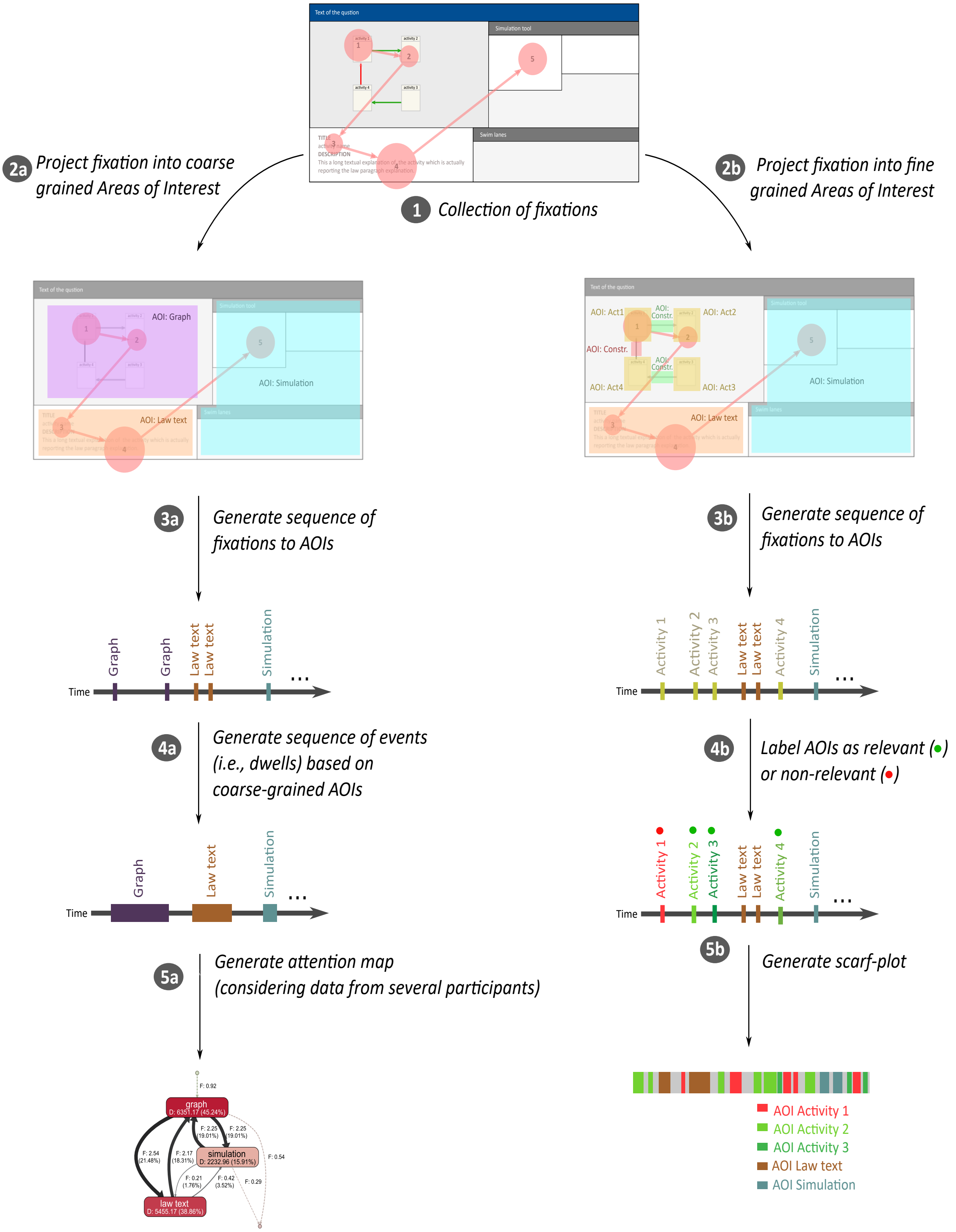
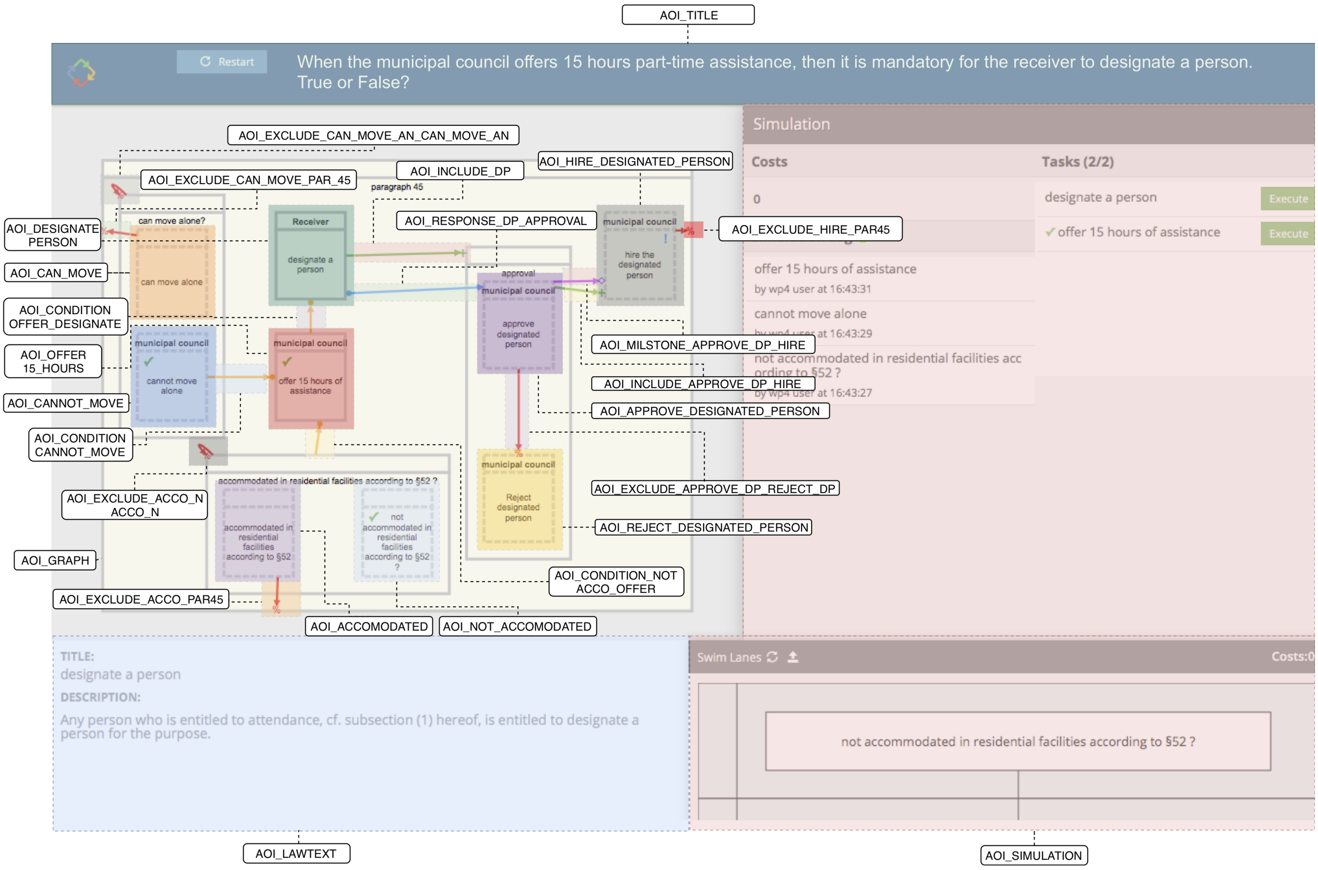


Fig. 3 Overview of the two eye-tracking data analysis approaches followed in this paper.



When the municipal council offers 15 hours part-time assistance, then it is mandatory for the receiver to designate a person. True or False?

TITLE:
designate a person

DESCRIPTION:
Any person who is entitled to attendance, cf. subsection (1) hereof, is entitled to designate a person for the purpose.

Fig. 4 Overview of the 25 areas of interest used for fine-grained eye-tracking data analysis. A higher resolution of this figure is available at `http://andaloussi.org/sosym2019/figures/DCR-HR-FineGrainedAOIs.pdf`

tention maps, the considered AOIs, corresponding to the three artifacts of DCR-HR, are represented as activities, whereas the transitions between them are represented as edges. Activities are labeled with the name of the artifact they represent, the mean dwell time (D), and the proportion of the overall time spent focusing on that artifact, which is represented as a percentage value and visualized through the intensity of the background color used for the activity. Relationships are labeled with the mean transition frequency (F) and with a percentage value quantifying the number of times that transition occurs compared to the total number of transitions. Graphically, this is rendered by the thickness of the arrows sketching the relationships. Finally, we represent the starting and ending points of the task and draw a dashed arrow connecting them to the first, respectively, last fixated artifacts.

In order to analyze the reading patterns of participants, we extracted the total fixation duration on each AOI (i.e., the dwell time [19]) and the frequency of transition between each pair of AOIs. Afterward, the mean fixation duration (D) and mean transition frequency (F) were derived by dividing each measure by the number of traces (i.e., sequences of dwells) used to discover the attention map. These two measures were projected respectively on activities and edges in the attention map to compare the reading patterns in different attention maps.

The second, more fine-grained analysis, followed to investigate **RQ3**, explores all the 25 defined AOIs and also considers the temporal dimension of fixation data, grounding on timeline AOI-based visualization approaches [23].

Specifically, after exporting the sequence of fixations (cf. step (3b) in Fig. 3) we examined how the obtained sequences of visits unfolded along a timeline skipping the aggregation into dwells done in the previous analysis (cf. step (4a) in Fig. 3). Indeed, this time we are interested in keeping a fine-grained resolution of fixations to know when and for how long a user fixated a certain AOI. This approach is beneficial to know when AOIs are visited during task execution and to detect consecutive revisits on the same AOI, keeping track of the saccades occurring between them.

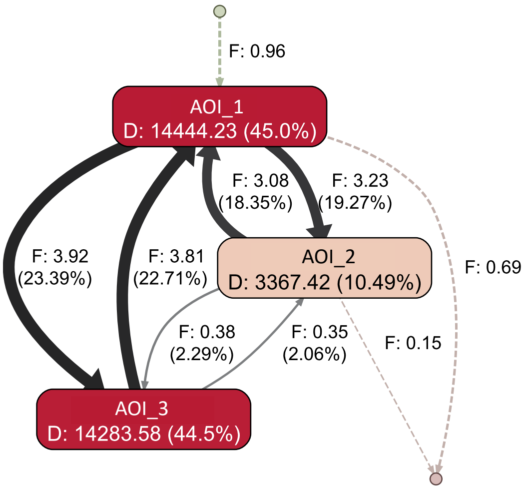


Fig. 5 Example of attention map showing three AOIs and the transitions among them. *D* represents the mean fixation duration while *F* is the mean transition frequency between two AOIs.

During this phase, we also labeled the AOIs of the DCR Graph based on whether they were deemed "relevant" for answering a particular question (cf. step (4b) in Fig. 3). Indeed, some comprehension tasks require the user to look for information that is explicitly represented in the DCR Graph and to combine it with contextual (law text) or execution (simulation) information to provide an answer. The AOIs related to the law text and the simulation have not been explicitly labeled as "relevant" or "non-relevant", but their relevance has been considered based on the kind of executed task.

Then, to visualize the visits on different AOIs over time we relied on scarf-plots [19, 59] (cf. step (5b) in Fig. 3), which show a timeline for each participant (i.e., a "scarf") divided into colored time spans. Each different color is used to encode a specific AOI, and the width of the time span is proportional to the duration of the fixation on that AOI. This visualization technique is particularly efficient for exploring and comparing scan-paths and sequences of dwells, as it shows which AOIs have been fixated often and when during task execution [23]. When creating the scarf-plots, we took inspiration from the qualitative alphabet palette of Polychrome, which includes 26 colors that are well separated in the CIE L*u*v* color space [63] and used all the colors in the palette apart from those in the range of greens. Indeed, a range of greens was created on purpose to be assigned to all the relevant AOIs in the graph to ease their identification during the analysis.

To create and visualize the scarf-plots used for the analysis, we adapted the piano-roll view of the Rythm-Eye tool [64]. Fig. 6 shows an example of scarf-plot depicting the sequences of fixations for three participants A, B and C. Each timeline captures the visual behavior of each participant over time. Colored time spans correspond to the visited AOIs: relevant AOIs are assigned a color belonging to the range of greens and are labeled in italic. It is worth noticing that the scarf-plot of Fig. 6 is normalized, that is, all the scarves have the same width regardless of the total time spent by each participant to execute the task. In this study, we relied on normalized scarf-plots as they facilitate the comparison of the visual behavior of different participants.

The analysis of scarf-plots was aimed to explore the users' visual behavior and to identify common strategies followed to engage with DCR-HR. To this end, we followed a qualitative coding approach. First, the scarf representing the behavior of each participant was inspected individually by appending qualitative memos [65] describing the observed behavior. During this process we relied on the distinction between goal-directed search and exploratory search introduced in the literature (cf. Section 2.3) to classify the participants' behaviors. We used two eye tracking measures (i) the time spent fixating AOIs that are "relevant" for executing the task (i.e., relevant areas of the DCR Graph and, possibly, the law text and the simulation based on the kind of executed task) and (ii) the total number of AOIs fixated during task execution. We assigned the label *goal-directed search* to the participants who focused on task-relevant information and visited only a small number of AOIs, while, we assigned the label *exploratory search* to the participants who switched their attention between a large number of AOIs, without showing particular interest for a specific AOI. We supported our labeling by triangulating the insights reported in the memos with the verbal data extracted from the think-aloud sessions, the eye gaze recordings, and some of the descriptive statistics inferred from the eye-tracking data. During this phase, co-authors dis-

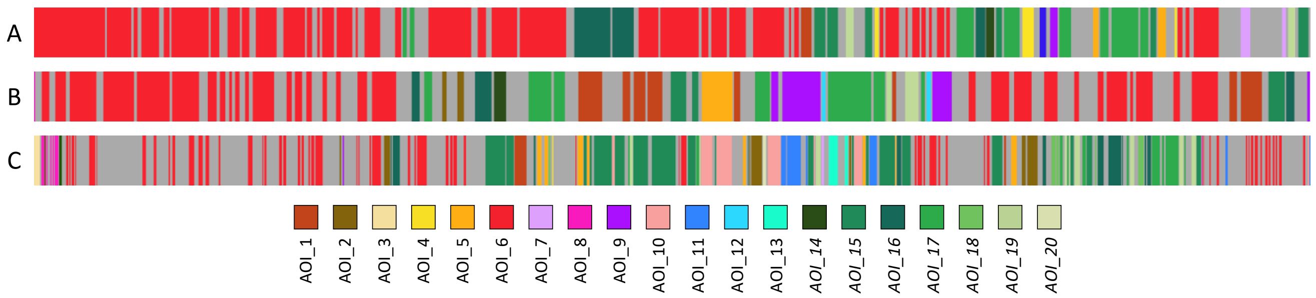


Fig. 6 Example of scarf-plot showing the fixations of three participants over time. Grey areas indicate that no AOI has been fixated at that particular moment.

cussed borderline cases to reach a consensus on how to classify participants exhibiting a potentially "mixed" behavior. Indeed, by going through gaze recordings, we noticed that even within the same task, some participants used more than one strategy when reading the law text or the DCR Graph.

Think-aloud data analysis. To answer **RQ2** we relied on think-aloud data. Prior to beginning with the analysis, we transcribed the verbal data collected during the retrospective think-aloud. Afterwards, we extracted the subjective insights from the transcripts with the support of Atlas.ti[8], a qualitative data analysis software tailored to deal with large bodies of textual, graphical, audio and video data [66]. During this process, we applied coding concepts from grounded theory [58], a methodology for developing theory by analyzing a text corpus, iteratively identifying recurring aspects and grouping them into categories. In particular, we used initial-coding to fragment the textual data and identify the emerging topics raised by the participants. Afterwards, we applied focus coding to identify recurring codes representing significant aspects into categories. Finally, we followed the principles of axial coding to find relationships among the inferred categories. The emerged codes serve as indicators of the use of the different DCR-HR artifacts and provide deeper insights about the patterns found in the eye-tracking data.

The complete analysis material can be found online[9]

5 Findings

This section reports the main findings of the exploratory study, presented according to the research question they concern. Section 5.1 presents the results related to the coarse-grained eye-tracking analysis. Section 5.2 reports the benefits and challenges associated with each of the DCR-HR artifacts. Finally, Section 5.3 describes the results of the fine-grained eye-tracking analysis.

5.1 How Do Users Engage with the Different DCR-HR Artifacts? (RQ1)

This section describes the results of analysis conducted to explore how people use DCR-HR. In particular, we explore whether the use of the different DCR-HR artifacts is influenced by (i) the participants' background or (ii) the kind of performed tasks.

(i) First of all, we begin with differentiating between municipal employees and academics based on their different backgrounds. Indeed, municipal employees are accustomed to reading legal texts but lack knowledge in process modeling, whereas academics have experience in process modeling but lack proficiency in legal reading. Based on their different backgrounds, we expect municipal employees and academics to prefer different artifacts (in line with their background) while performing the comprehension tasks.

To explore this assumption, we used both attention maps generated from the eye-tracking data and aggregated over all participants and the insights derived from the think-aloud data. Fig. 7a and Fig. 7b show the attention maps comparing the reading patterns of municipal employees and academics when answering the given tasks.

By taking a closer look at them, we can see that both groups started by observing the DCR Graph, which is reasonable since the DCR Graph is placed in the center of the screen and occupies a large portion of it (cf. Fig. 1). Furthermore, we can observe that academics spent substantially more time looking at the different artifacts compared to municipal employees (cf. mean fixation duration D in Fig. 7a and Fig. 7b). This observation is also supported by the subjective insights retrieved from the think-aloud transcripts. Indeed, municipal employees reported to have relied on common

---

[8] Atlas.ti, a qualitative data analysis tool. See `https://atlasti.com`

[9] see `http://andaloussi.org/sosym2019/analysis/`

sense or on their working experience when answering to some tasks (e.g., *"If the recipient is unsatisfied, then, of course, you can change the decision* [while the DCR Graph shows clearly that such a decision cannot be reversed]*"*[10]). Instead, academics claimed to have never relied on common sense when performing the tasks.

The attention map depicted in Fig. 7a reveals that most municipal employees split their attention mainly between the DCR Graph and the law text. From the think-aloud, we found evidence that the majority of municipal employees did not use the simulator, but relied only on the graph and the law text (e.g., *"I have either read through the law text or the model but I have not used the simulation."* [10]). However, some municipal employees used all the three artifacts. In particular, a municipal employee affirmed to have used the law text but having relied on the simulation to validate his answers, while another municipal employee mentioned using the simulation twice during the whole experiment. These insights line up also with the proportions of transitions between the artifacts. Indeed, Fig. 7a shows that the number of transitions between the graph and the simulation and between the graph and the law text is similar, suggesting that municipal employees have generally interacted with all the different artifacts.

When looking at the attention map of Fig. 7b, outlining the use of artifacts carried out by academics, we can observe that artifacts have been used differently. Academics focused mostly on the DCR Graph and split the rest of their attention between the law text and the simulation. Accordingly, the proportion of transitions between the different artifacts shows that academics did almost twice more transitions between the graph and the simulation than between the graph and the law text. Indeed, academics not only spent a limited time on the law text compared to municipal employees but have switched less often between the DCR Graph and the law text. By looking into think-aloud data we had confirmation that many of the academics struggled to understand the legal terms and the linguistic patterns used in the law text (e.g., *"I think understanding this law jargon was kind of difficult"*, *"I tried to read the law text to understand the law but it actually didn't help at all because the language that is used is pretty formal"*) and, therefore, may not have found it useful.

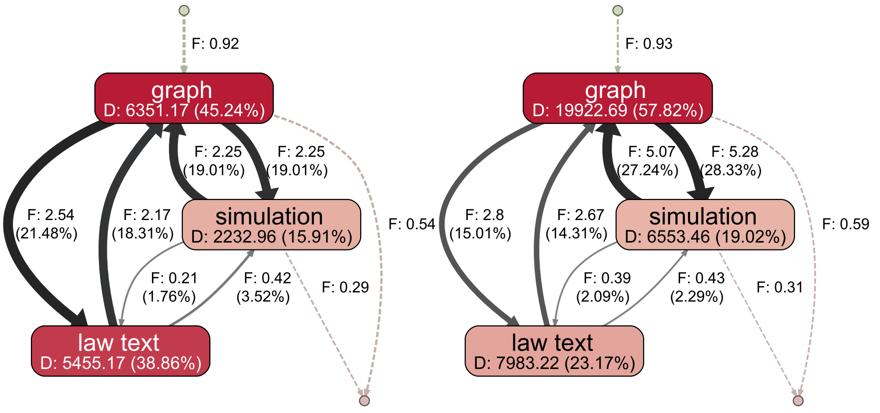
Overall, the attention maps shown in Fig. 7 suggest that users with different backgrounds would use different artifacts to understand the process. Hereby, the hybrid nature of DCR-HR can provide a unified representation that can make process models accessible for users with different backgrounds.

(ii) Besides looking into how the users' background affects the use of different artifacts, we investigated whether the choice of which artifacts are used changes when dealing with different tasks. To explore whether constraint, decision, and scenario tasks (cf. Section 4.2) are executed with the help of different artifacts, we rely again on attention maps.

Fig. 8 depicts the attention maps summarizing how all participants used the artifacts of DCR-HR when executing constraint, decision and scenario tasks respectively. These visualizations reveal a different use of artifacts for each type of task. Fig. 8a shows that in constraint tasks (i.e., questions asking to focus on the relationships between pairs of activities of the DCR Graph), the participants focused mostly on the DCR Graph, and split the rest of their attention between the simulation and the law text. This use of the graph can be explained by the nature of constraint tasks and it is confirmed by some subjective insights obtained from think-aloud data. However, some other participants seemed to be challenged with the semantics of DCR relations and were often resorting to the simulation to clarify the implications that the different relations have on the model behavior (e.g., *"The simulator, I used it when I was in doubt because, the different arrows I wasn't always sure what they did, so then I rendered simulator . . . then you could actually know for sure if you could do this after this or not"*). These subjective insights find confirmation in the high number of transitions between the DCR Graph and the simulation. Indeed, the participants did twice as many transitions between these two artifacts than between the graph and the law text.

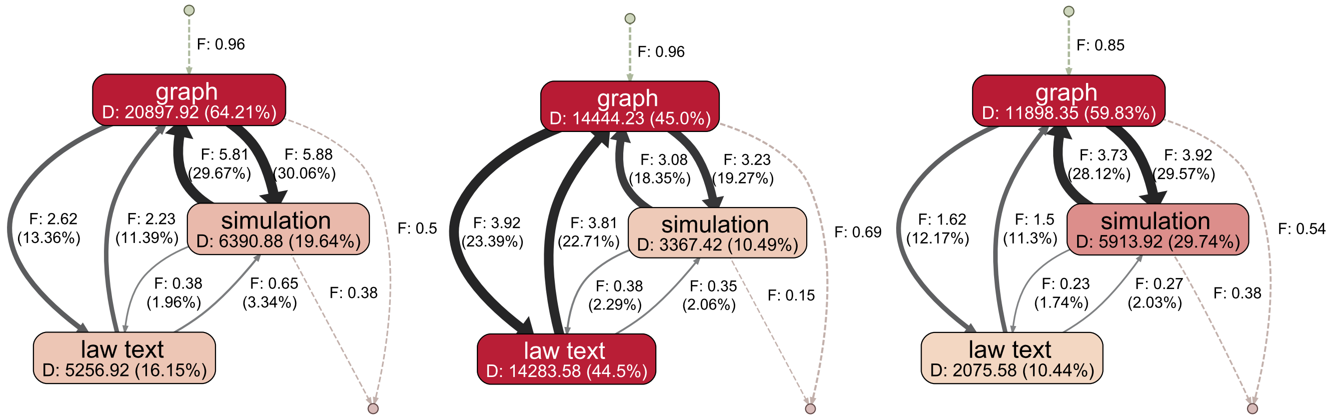
When examining the attention map related to decision questions, shown in Fig. 8b, we can observe that participants split their attention mainly between the DCR Graph and the law text. This evidence suggests that, when asked to choose among multiple options, participants relied on the law text to retrieve contextual information useful to support their decision-making process. This explains also the high number of transitions between the graph and the law text occurring in decision tasks.

Finally, Fig. 8c shows how the different artifacts have been employed in scenario tasks. By looking at the attention map, we can clearly see that the participants spent relatively less time on the law text while switching their attention mainly between the DCR Graph and the simulation. Hereby, one can argue that the participants have mainly combined the DCR Graph and the simulation to answer scenario tasks. This assumption is supported by the think-aloud data where participants affirmed using both the DCR Graph and the simulation when asked to determine the behavior of the process model (e.g., *"When the question is in a scenario then I*



(a) Municipal employees (b) Academics

Fig. 7 Attention maps comparing the attention focus on different artifacts for municipal employees and academics. *D* is the mean fixation duration and *F* is the mean transition frequency between two AOIs.



(a) Constraint tasks (b) Decision tasks (c) Scenario tasks

Fig. 8 Attention maps showing the use of different artifacts for all participants executing different types of tasks. *D* is the mean fixation duration and *F* refers is the mean transition frequency between two AOIs.

use the simulator, because it's easy to see what happens after").

The attention maps for constraint, decision and scenario tasks share one common trait that is the limited number of transitions between the law text and the simulation. The collected think-aloud data support this evidence, as participants could not find any circumstance where the combination of law text and simulation was beneficial (e.g., *"Simulation and law text doesn't go well together because that if you can actually solve it with the simulator, you don't need the law text"*).

Overall, the insights outlined in Fig. 8 show that the participants combined different artifacts when executing different tasks. On this ground, we argue that the deployment of hybrid process artifacts such as DCR-HR can support users dealing with different tasks in a situation-specific manner.

5.2 What Are the Benefits and Challenges Associated with Each One of the Artifacts of DCR-HR? (RQ2)

In this section, we rely on think-aloud data to gather insights into the perceived benefits and challenges of DCR-HR artifacts and to improve our understanding of the use behaviors discussed in Section 5.1.

The results of the think-aloud data analysis show that the DCR Graphs helped several participants to get a good overview about the business process (e.g., *"The model* [...] *I mainly used it to identify how the overall process works"*). The DCR Graph was also used by participants to identify and navigate through the law text (e.g., *"You can highlight different sections of* [the] *law through* [the] *model"*). Some academics reported that the DCR Graph helped them to understand the interplay between the different process activities (e.g., *"I use the model to see* [the] *interaction between the four dif-*

Artifact	Benefit	Challenge
DCR Graph	• Provides a good overview of the process • Helps navigating the law text	• The semantics is hard to understand (**ME**) • The model does not capture enough details
Law Text	• Provides more information than the graph • Supports decision-making (**ME**)	• Hard to read (**AC**)
Simulation	• Clarifies the DCR semantics • Allows testing different executions • Can be used to validate assumptions made on the graph	• Inefficient for evaluating all the possible process executions (**AC**) • Time-consuming (**ME**)

Table 2 Challenges and benefits resulting from the analysis of the think-aloud data. The labels **AC** and **ME** highlight whether a certain benefit/challenge was perceived only by academics (i.e., **AC**) or municipal employees (i.e., **ME**). All the other benefits and challenges are perceived by participants belonging to both groups.

ferent activities"), whereas some of the municipal employees seemed to find DCR relations hard to understand and use. These challenges were inferred from the think-aloud, as several participants reported difficulties in identifying the appropriate DCR relation specifying a certain behavior. Last but not least, some participants found the DCR Graph very abstract and pointed out that the model was not capturing all the details specified in the law text (e.g., *"If you only have the model it's very abstract"* or *"The strange thing is that many things which the law is talking about the model did not talk about"*).

The legal text, in turn, provided participants with details which were missing in the DCR Graph (e.g., *"I guess it provided more details in some cases than the model"*, *"The law text might be able to add some details that can't be in the model "*, *"If I didn't think that the model accurately captured enough for me to answer the question, then I would read the whole text instead"*). The participants also mentioned that the law text was effective to support their decision-making process when the DCR Graph allowed for more than one choice (e.g., *"When I had to use the law text, it was for questions about -should I do this at all- for example should I give personal permission, should I take the accept or should I take the reject button on an activity."*).

Several municipal employees expressed a preference for the legal text as they were already familiar with reading and interpreting law paragraphs (e.g., *"I mostly used the law text because that's what I'm used to look at"* [10]). In turn, many of the academics struggled to understand legal terms (e.g., *"I think understanding this law jargon was kind of difficult"*, *"I tried to read the law text to understand the law but it actually didn't help at all because the language that is used is pretty formal"*). For this reason, some academics did not use the law text to gather information about the business process (e.g., *"It is not so easy to read the law text ... I have totally ignored it"*).

Finally the interactive simulation allowed the participants to check the viability of different process executions (e.g., *"The simulation is helpful to see the possible paths"*, *"You can actually see if you have a viable execution"*). Moreover, some academics affirmed that using the simulator helped to reduce the mental effort required to keep track of the all dependencies existing among different DCR relations (e.g., *"It's a little too much to have all the steps in your mind while you're going ... "*, *"It is easier to see it simulated instead of manually analyze the model"*). These comments fall in line with the previous claims about the role of interactive simulations in improving the understandability of declarative process models [4].

The analysis of the transcripts shows also that the simulation helped participants to validate the assumptions they had made by looking first at the DCR Graph (e.g., *"You can like simulate the process then you like get a clear understanding of how the process works ... if you're in doubt of like relations or anything in the graph then you can use the simulation to like confirm what you actually think about the model "*], *"... checking if it is exactly what I thought the model is doing it's actually doing it"*). Yet, other participants pointed out a few drawbacks associated with the use of the simulation. In particular, some academics considered it inefficient having to restart the simulation every time an undesired state was reached (e.g., *"Actually this was not very convenient because you click the all way through and if you miss a click, which I actually did, you need to do it again"*). Others abstained from using the simulation because they were able to mentally simulate the execution of the process (e.g., *"Primarily, I didn't use the simulator at all because I pretty much simulated in my head"*). Municipal employees, who are not used to interact with simulation interfaces on a daily basis, perceived the simulation as time-consuming (e.g., *"I'm*

---

[10] Quote translated from Danish.

used to work under very high work pressure, so getting in and checking such things through that way is not in my habits" [10], *"You would spend too long to press and read all four options, then press again and read three new options, then press again and there will be five new options"* [10]).

The analysis conducted in this section shows that each artifact has some strengths but also presents some weaknesses, which are summarized in Table 2. Despite having been exposed to the three process artifacts during all the comprehension tasks, participants showed a preference for certain process artifacts based on their domain knowledge, the perceived usefulness and the context in which the artifacts have been deployed. Participants have also reported a set of challenges they faced when interacting with these artifacts.

The perceived benefits and challenges identified by users are often based on their background and working habits and, therefore, are likely to influence the use of the different artifacts. For this reason, benefits and challenges are also suitable to complement the findings of the analysis introduced in Section 5.1 and to provide a deeper explanation of certain use behaviors. For example, from the analysis of the attention maps of Fig. 7a we observed that the simulation was used rarely by municipal employees, especially when compared with the other artifacts. As Table 2 reports, the think-aloud data reveal that municipal employees considered it as "time-consuming" for a person that is not used to deal with such tools in his or her working life. Similarly, the inherent complexity of the legal text seems to have discouraged academics to use it, as reflected in attention map of Fig. 7b.

Overall, the results presented in this section suggest that participants with different backgrounds perceive different benefits from the use of different artifacts. Thereby, we can claim that combining all these artifacts into a hybrid representation makes their use more effective and eases the task of meeting individual user preferences.

5.3 What Strategies Are Followed When Engaging with the Different DCR-HR Artifacts? (RQ3)

By analyzing the data following the procedure outlined in Section 4.3 we were able to classify the behavior of participants into goal-directed and exploratory. We noticed that participants exhibiting a goal-directed behavior fixated mainly the question title and the AOIs in the DCR graph deemed relevant for a particular task. Moreover, we observed that fixations on the law text often occurred when solving decision tasks (i.e., when contextual information is needed) while fixations on the simulation often occurred when solving scenario tasks (i.e., when information about execution traces is required). In addition, we noticed that the relevant AOIs of the graph were either visited following a sequential pattern, that is by visiting one relevant AOI after the other, or following a more fragmented pattern, that is by switching frequently between the question title and the relevant AOIs. As for the participants who exhibited an exploratory behavior, we observed that they fixated more AOIs in the DCR Graph. However, these AOIs were often irrelevant for answering the given tasks. Moreover, we noticed that these participants were continuously intertwining between the different artifacts of DCR-HR and within different parts of the DCR Graph without focusing on a particular AOI, which in turn hints towards a lack of a guided strategy to solve the given tasks.

Fig. 9 illustrates the visual behavior (represented as sequences of fixations projected on the stimulus and scarf-plots) of two representative participants solving a constraint task: participant P09 following a goal-directed strategy and P14 exhibiting an exploratory strategy.

The considered constraint task was asking the following question: "When the municipal council offers 15 hours part-time assistance, then it is mandatory for the receiver to designate a person. True or False?". Clearly, since the question mentions information that is explicitly represented in the DCR Graph, we identified activities *offer 15 hours of assistance* and *designate a person* as relevant, together with the condition connecting them (cf. Fig. 1). In Fig. 4 these correspond to the areas of interest labeled AOI_OFFER_15_HOURS, AOI_DESIGNATE_PERSON and CONDITION_OFFER_DESIGNATE. All the other areas of the graph are considered non-relevant for this specific task, whereas the law text and the simulation could in principle being used to support the user in responding to the question.

As regards to the number of visited AOIs, P09 visited only 5 out of the 22 AOIs defined on the DCR Graph and most of the fixations were on relevant AOIs. In addition, a large portion of these fixations lasted for a considerable period of time: P09 spent the 35.31% of the total fixation time on relevant AOIs, while the same proportion for non-relevant AOIs of the graph amounts to 3.74%. By considering the temporal order of the fixations, it is easy to see that P09 had several long and repeated fixations on the question title at the beginning of the comprehension task. Afterwards, the participant visited AOI_OFFER_15_HOURS and AOI_DESIGNATE_PERSON in sequential order for some time before switching back to the question title.

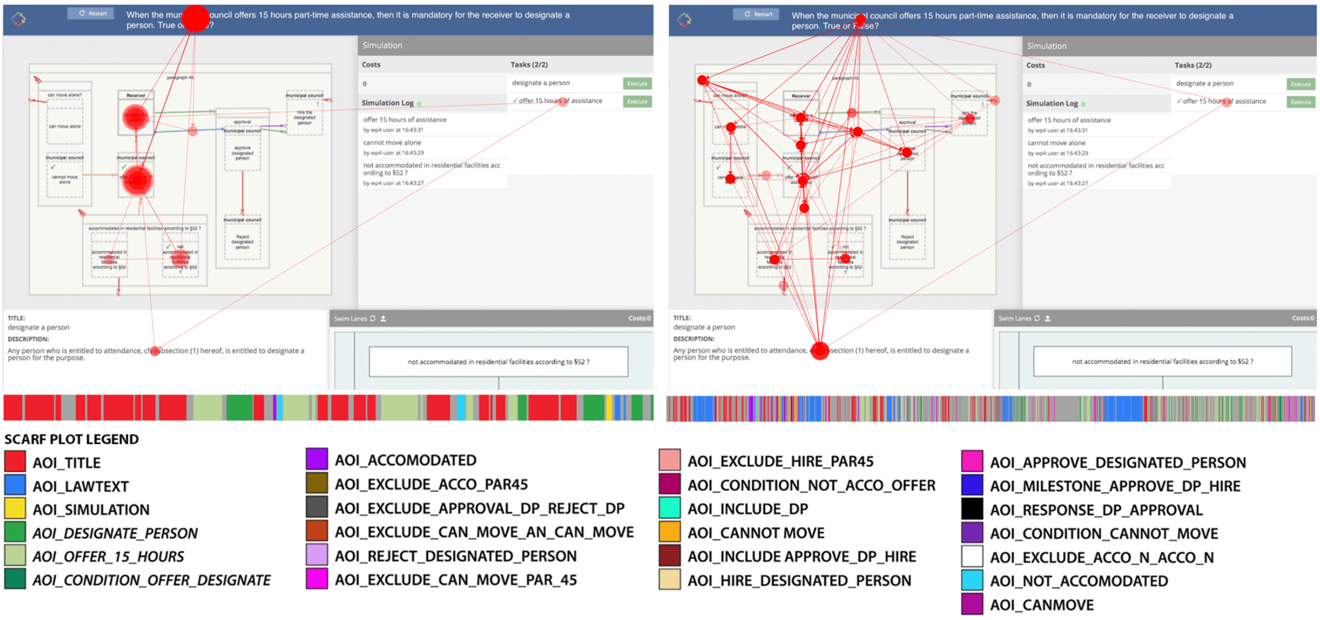


Fig. 9 Comparing the visual behavior of participant P09 following a goal-directed strategy to solve a constraint task (left) to the one of participant P14 following an exploratory strategy (right).

The same pattern re-occurred several times throughout the whole comprehension task. In addition, P09 did not focus on artifacts other than the DCR Graph. This suggests that the user is familiar enough with the semantics of DCR Graph and aware of how to derive the needed information from it. We labeled the strategy followed by P09 as "goal-directed".

By taking a look at the behavior of P14, we can see that this participant visited several AOIs during the execution of the task and, precisely, he focused on 16 out of the 22 AOIs covering the DCR Graph. A large amount of these fixations were on non-relevant AOIs and most of them lasted for a short amount of time. Overall, P14 spent the 29.08% of the total fixation time fixating relevant AOIs in the graph, and the 29.33% of it fixating non-relevant ones. By considering the temporal order in which the fixations occurred, one can see that P14 had several short fixations on the title before moving briefly to the law text and then coming back to the title. Afterwards, the participant kept intertwining between different AOIs apparently without focusing a particular one. This pattern continued throughout the whole comprehension task thus suggesting an exploratory strategy.

Indeed, during the retrospective think-aloud, P14 reported some difficulties with remembering all the different kinds of constraints included in the DCR Graph and he admitted having used the law text to improve his own understanding of the different relationships: *"I did remember from the introduction I was given . . . I did remember a few things but through the time I was forgetting a bit what do they mean to me so I was using the hold over* [the law text] *to understand them to see what they mean"*, *"the logic behind them* [the constraints] *was clear was clear but it's a matter of remembering"*.

The goal-directed and exploratory behaviors illustrated for P09 and P14 can be observed in all participants, regardless of the kind task they have been solving. Fig. 10 shows the scarf-plot of all participants related to the constraint task discussed for P09 and P14. As observed for the latter ones, the difference between the goal-directed and exploratory strategies does not lie in the amount of time spent on the relevant AOIs of the graph which (19,97% vs 24,67%), but is rather noticeable in the amount of time spent on non-relevant AOIs (12,83% vs 27,04%). Moreover, we observed that the total task execution time for participants exhibiting a goal-directed behavior was shorter on average compared to the one of those showing an exploratory behavior: the first group took an average of 44 seconds, while the second group took 02 minutes and 18 seconds on average. During the execution of this task, participant P03 exhibited an unusual behavior, as he fixated mostly the title and had only a couple of glimpses on non-relevant AOIs of the DCR Graph. According to [67], long fixations on a misleading element seem indicate an unclear interaction behavior. When looking into the verbal comments gathered with the think-aloud we discovered that P03 was confused about the graph and used common sense to respond to this question, which explains the unusual strategy.

Fig. 11 and Fig. 12 show the visual behavior of the participants solving a decision task and a scenario task, respectively. Considerations similar to those made for Fig. 10 can be made regarding the categorization of users' strategies into goal-directed and exploratory, keeping in mind that for answering the decision question resulting in the plots of Fig. 11 users had to necessarily refer to contextual information included in the law text, while for answering the question resulting in the plots of Fig. 12 the simulation would have made it easier to remember the process trace.

Regardless of the kind of question being solved, we observed other patterns in the use of different artifacts over time. In particular, we noticed that simulation (coded in yellow in Fig. 10 – Fig. 12) was used at different times during task execution, that is, either immediately after having read the question title (as done by P06 in Fig. 11 or by P05 in Fig. 10 and in Fig. 12) or towards the end of the task, as done by (P02, P06, P11 and P13 in Fig. 12).

This, in turn, suggests two different ways of using simulation: people using it at the beginning of the task may have exploited it as a support for understanding the question and to improve the interpretation of the graph, while people using it towards the end of the task may have used it to validate or confirm a hypothesis they had made by looking at the graph or the law text.

By looking into the think-aloud, we found evidence of these two different uses of the simulation. Participant P06 seems to confirm that simulation is useful to understand what the question is asking, while P11 reports that the simulation eases the comprehension of the different relations in the graph.

More precisely, P06 mentions *"I think I used a lot the simulator. It is useful. Basically, you just base on the question and you follow the question and you do it"*. *"For example in a question there will be kind of assumptions or kind of simulation about if you are alone or with your mom. I used the simulator and I said - Ok, what is the possibility and what are the next steps - then I would say - Ok back to the text* [i.e., the question]*... what he wants me to do? He wants me to go for this possibility, then I click this one - so I can align the text and the simulator."* P11 reports *"I used the simulator and then I learned some things that I wasn't sure about. Also if I was sure about what would happened in one sub-process and then I saw it, then next time I knew how it worked"*.

Other participants confirmed having used the simulation as a way to confirm something they were not sure about or something they had already hypothesized when looking at the graph. Participant P05 claims to have used the simulator as an exclusion criterion: *"Those* [the questions for which he used the simulator] *were the ones where there were more criteria put into the question. Then I used it as a method of exclusion, I think."*. Again, participant P11 reports having used the simulator to confirm the answers he had in mind after having looked at the model *"I use it* [the simulator] *after... if I use the model and I'm not sure or maybe I just want to convince myself. Like, maybe I think this is the right answer but just to make sure I can run the simulator because it kinds of makes me surer."*.

Last but not least, considering the sequential order in which the tasks were displayed to the participants and the strategies adopted when executing different kinds of tasks, we noticed a trend towards switching to a more goal-directed strategy as each eye-tracking session proceeded. Table 3 summarizes the kind of behavior adopted by each participant when solving the constraint task (cf. Fig. 10), a decision task (cf. Fig. 11) and a scenario task (cf. Fig. 12). Hereafter, one could notice that some of the participants who started with an exploratory behavior switched to a goal-directed one during the eye-tracking sessions.

Participant P11 provided a possible explanation for this finding in the think-aloud: *"The first round I spent a lot of time looking at the whole model even though it didn't have anything to do with the questions because I wasn't sure if I just missed something and then I could use that information later on."*. This kind of behavior seems to find confirmation in the fact that exploratory search can sometimes operate as a screening process that identifies candidates for goal-directed search [24].

6 Discussion

In this section, we discuss the findings of this study in light of existing literature. Sections 6.1, 6.2 and 6.3 discuss the findings associated with **RQ1**, **RQ2**, and **RQ3** respectively, while Section 6.4 discusses the limitations of this work.

6.1 How Do Users Engage with the Different DCR-HR Artifacts? (RQ1)

The results presented in Section 5.1 suggest that people interact differently with the artifacts of DCR-HR based on their background and the kind of task they are executing. The way different backgrounds affect the use of DCR-HR is in line with the claims made by previous researchers [2, 3, 44] to support the purposeful perspectives offered by hybrid process artifacts. Moreover, our findings reflect the circumstance where

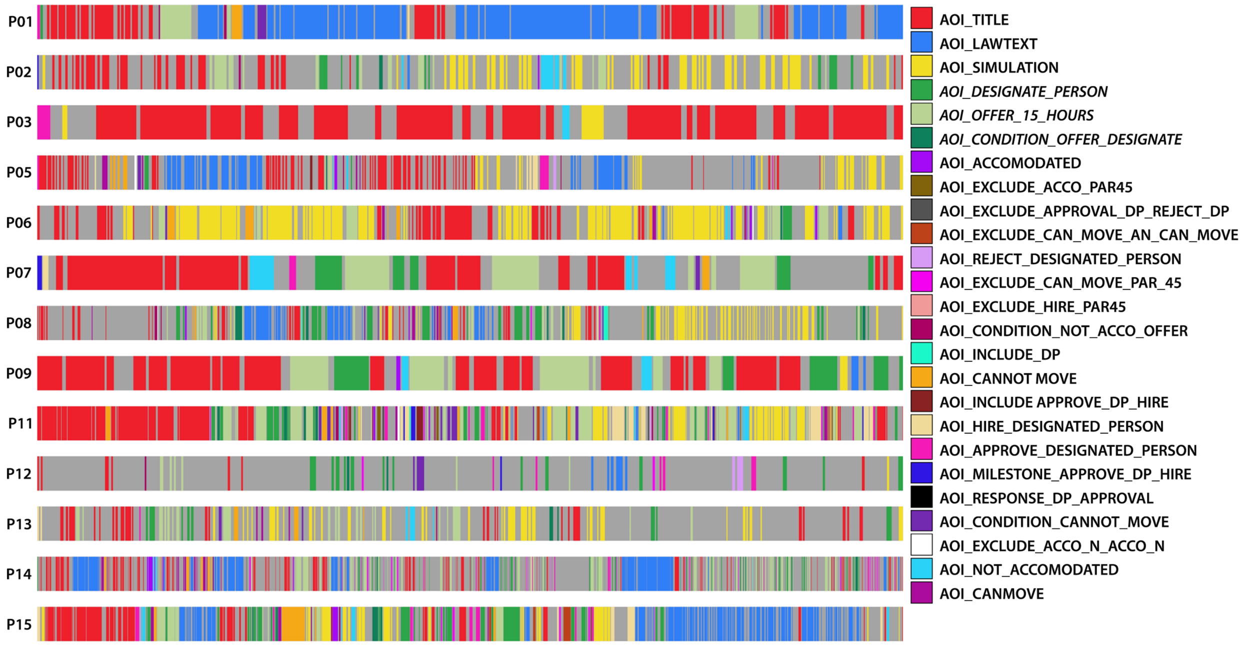


Fig. 10 Scarf-plot showing the sequences of fixations for participants solving a constraint task and classified into goal-directed (GB) and exploratory (EB) strategies. Relevant AOIs of the DCR Graph for this task are labeled in italic.

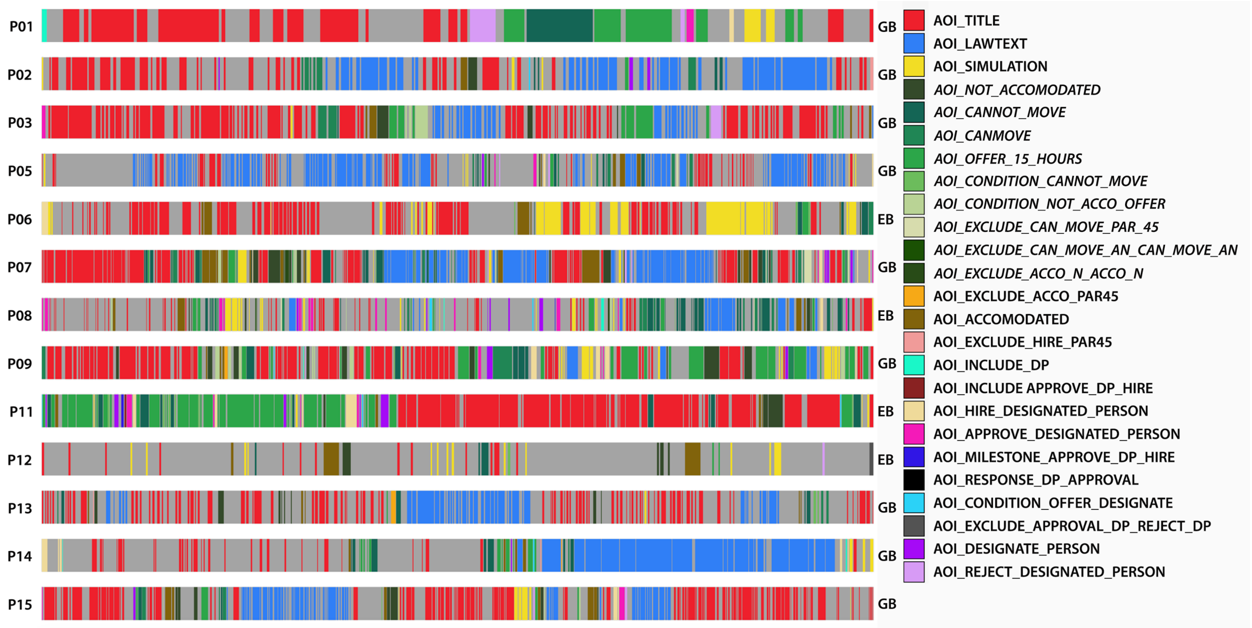


Fig. 11 Scarf-plot showing the sequences of fixations for participants solving a decision task and classified into goal-directed (GB) and exploratory (EB) strategies. Relevant AOIs of the DCR Graph for this task are labeled in italic.

both domain experts (represented by municipal employees) and IT specialists (represented by academics) are challenged when being exposed to unfamiliar process artifacts [3]. The deployment of a hybrid process artifact can help to overcome this issue by providing a representation that is comprehensible to both stakeholders.

Existing research associates the comprehension of business processes with the effectiveness of communication between different stakeholders [3,68]. Using different levels of formality (i.e., natural language and DCR notation) and different levels of abstraction (i.e., a process model abstraction and an instance-based simulation) hybrid process artifacts could foster the communication between different groups of stakeholders by providing the means to clarify the terms and relationships in the domain and prevent misinterpretations.

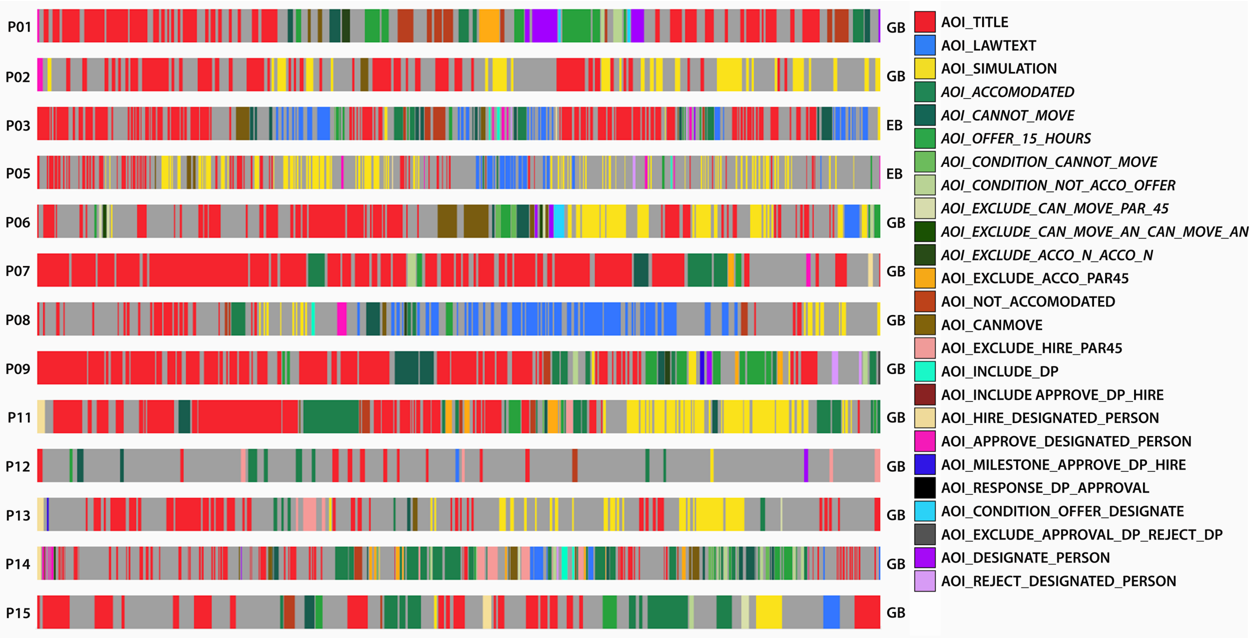


Fig. 12 Scarf-plot showing the sequences of fixations for participants solving a scenario task and classified into goal-directed (GB) and exploratory (EB) strategies. Relevant AOIs of the DCR Graph for this task are labeled in italic.

													TOTAL	
Task type	P01	P02	P03	P05	P06	P07	P08	P09	P11	P12	P13	P14	P15	**GB**	**EB**
Constraint	GB	GB	–	EB	GB	GB	EB	GB	EB	GB	GB	EB	EB	**7**	**5**
Decision	GB	GB	GB	GB	EB	GB	EB	GB	EB	EB	GB	GB	GB	**9**	**4**
Scenario	GB	GB	EB	EB	GB	GB	GB	GB	GB	GB	GB	GB	GB	**11**	**2**

Table 3 Summary of goal-directed and exploratory strategies across tasks. GB refers to Goal-directed Behavior; EB refers to Exploratory Behavior.

The disparity of task types is clearly reflected in the reading patterns of the participants. From the analysis conducted in Section 5.2, it has emerged that the participants have changed their reading patterns according to the task they were executing. This suggests that different combinations of artifacts have been used in a context-specific manner to address different comprehension tasks.

The relation between task types and the underlying cognitive processes are discussed in the literature in a bi-directional fashion. In cognitive psychology, the impact of the task type on the underlying cognitive processes is discussed by Glaholt et al. [69]. Following a quantitative approach, the authors investigate the eye gaze selectively (i.e., the extent to which an individual is selective when being asked to choose between different alternatives) associated with different types of tasks. They discovered that the task type has a direct influence on eye-tracking measures, such as the total fixation duration, number of dwells, and their mean duration. Grounded on a set of visual patterns, our attention maps come to support Glaholt's insights from a qualitative perspective showing the difference in users' reading patterns when approaching different model comprehension tasks.

In the field of process modeling, existing research shows the potential of deploying users' behavioral patterns to predict the task they are involved in (i.e., problem understanding, method finding, validation) [70]. Interestingly, many of the measures deployed in [70] (i.e., dwells and transitions between AOIs) to provide accurate predictions about the task at hand, have also been used in the present study as a basis to generate the attention maps showing the distinctive reading patterns. Hence, the emerging distinction of task types could be operationalized using the underlying measures to provide context-adaptive tool-support at run-time, making relevant aspects more salient and accessible for the user.

6.2 What Are the Benefits and Challenges Associated with Each One of the Artifacts of DCR-HR? (RQ2)

The retrospective think-aloud provided deeper insights into the way users engaged with the different DCR-HR artifacts. These subjective insights provide a means to explain the different reading patterns inferred from the eye-tracking data – but also to identify the challenges associated with the different DCR-HR artifacts. The challenges associated with the comprehension of the DCR Graphs intersect to a large extent with the notorious challenges of declarative process models [5, 28–30]. When it comes to the preference for textual languages (e.g., law text) over graphical ones (e.g., DCR Graphs) or the other way around, the background of the participants seems to play a central role. A similar finding is reported by Ottensooser et al. [71]. The authors associate the experience of the user with a particular language to the effectiveness of its use. This is also supported by other literature (e.g., [72, 73]) emphasizing the importance of individuals' familiarity with the language being used and its impact on creating a better cognitive fit between the given material and the task in hand. As a result, this insight could explain the preference of municipal employees to law text and academics to DCR-graphs. Besides this personal factor inherent to the participants' background, the existing literature remains divided on whether graphical languages are more comprehensible than textual ones. Throughout different studies, authors compare several textual and graphical languages, and the results are split based on the languages being compared [71, 74, 75]. One the one hand, graphical languages for process modeling can be efficient with regards to their spatial arrangement of the modeling constructs, allowing the reader to perceive the interplay between the different constructs easily [76]. However, the deployed graphical notations are not as intuitive as they claim to be [77] and usually require formal training to be understood [78]. On the other hand, textual languages can be readily comprehended and interpreted by non-experts [71,78]. The exploratory insights of this study suggest that this argument could only hold for readers having a precedent textual aptitude for a specific domain terminology; otherwise, the deployed vocabulary would still remain challenging to be understood. This is exactly the case of academics, who were mostly challenged by reading the law text. The deployment of a hybrid process artifact could, in turn, help overcoming the challenges associated with the individual artifacts. In cognitive psychology, this suggestion is largely supported by the dual coding theory [79], which highlights the importance of combining textual and graphical artifacts to reinforce the understandability of the material in hand.

Looking further into the subjective insights provided by the participants, it seems that academics and municipal employees perceived different precision and rigidity in the DCR Graph and the law text. These perceptions can be seen as being subjective to the participants' background. Indeed, academics did not feel comfortable with the fuzzy law text and showed a preference for the process model, being a formal artifact prescribing the exact behavior of the model. In contrast, municipal employees rather used the law text, which is less formal and leaves room for diverse interpretations. While it is true that a formal model is more precise and less ambiguous than law text, it might not be sufficient for solving all possible cases. As the level of flexibility supported by the model is predetermined at design time, a model could either encode a strict interpretation of the law, or a loose one, but never both. The law text, in contrast, can support both flexible and strict interpretations depending on the way the reader perceives it. Another possible limitation of the model is due to its inability to specify the most desired path and outcome of the process, whereas the law text could explicitly express this requirement.

6.3 What Strategies Are Followed When Engaging with the Different DCR-HR Artifacts? (RQ3)

Among the findings of this study is the mapping between the different strategies enacted by participants and the visual search behaviors (i.e., goal-directed, exploratory) introduced in the literature (cf. Section 2.3). Metrics such as the distribution of fixations over time, their frequency and duration, and the number of fixated AOIs were previously used in quantitative studies to distinguish different types of visual search behaviors [25] or to identify usability issues [67]. In this paper, we followed a qualitative approach supported by eye-tracking to map the strategies enacted by participants to the existing visual search behaviors and backed up the interpretation of our data with insights derived from the think-aloud.

The time spent fixating non-relevant AOIs was among the key features used for this classification. According to [46] the time spent fixating relevant regions of a process model can predict answer correctness. In our study, we did not notice such correlation between the time spent on relevant AOIs and the amount of correct answers. However, while the authors in [46] consider BPMN process models and focus on structure understanding, we focus on general comprehension questions on DCR Graphs, whose declarative nature requires people to comprehend the different constraints involved and hidden dependencies [53]. In this context, even if a relevant AOI is fixated for a long time, a misunderstanding of the constraints semantics may compromise the correctness of the provided answers. Moreover, especially for decision tasks, looking solely at relevant AOIs in the process model was not enough to come up with a correct answer.

In usability studies, short fixations targeting a large number of non-relevant AOIs have been associated to confusion generated from the inability of the user to find a certain piece of information [67]. In the think-aloud, confusion was indeed one among the most recurring issues reported by participants, mostly in relation with the DCR Graph or the law text. Yet, some people claimed that their understanding of the information conveyed by the three artifacts improved throughout the study, thus allowing them to make a more effective use of each artifact and resolving part of the initial confusion. This behavior could explain why we noticed a change from exploratory to goal-directed strategies over time. Such change of behavior across eye-tracking sessions is also reflected in the findings of a recent experiment conducted by Zhao et al. [39] in the context of text-picture comprehension. According to the authors, pictures are likely to be used for task-oriented specific processing, but participants require some time to construct their own mental model before using them. Although DCR Graphs are not simple pictures as they include textual labels and a graphical notation having a precise semantics, we found evidence of such behavior in the think-aloud data (cf. Section 5.3) and, therefore, we may consider the findings in [39] as a possible explanation for the change in strategies.

Last but not least, we were able to observe also other patterns describing the use of the different DCR-HR artifacts and, specifically, the use of simulation. Declarative process models such as DCR Graphs are known to be well-suited to convey circumstantial information, while sequential information (e.g., traces) remains implicit [29]. Thus, to determine whether a particular trace is supported or not, or to determine possible following actions based on a partial trace, the user is required to keep numerous states in mind. The simulation, in turn, makes the sequential information explicit and allows to offload memory facilitating the execution of scenario tasks. In this study, we noticed that some participants used simulation starting from the initial phases of the comprehension task as a means to improve their understanding of the question or of the DCR Graph. Instead, others used the simulation towards the end of the comprehension task, to sort of validate their hypothesized answers. Pinggera highlights a similar use of validation during process modeling. In his PhD work [55], he reports about the difference between incremental and final validation. In the first case, validation, co–occurs with the modeling and reconciliation phases, indicating that the modeler spent some time evaluating the designed process model. In the second case and most common scenario, validation is observed mostly at the end and the time spent validating the model seems to increase according to the size and the complexity of the model.

6.4 Threats to Validity

There are number of threats associated with the validity of our research. In the following we discuss these threats and evoke the actions taken to mitigate their effects.

Internal Validity. The design of the study is subject to some threats. Our study uses a relatively small sample size, which is nonetheless acceptable in exploratory studies [80,81]. Moreover, disparities in the expertise of the participants might have limited the interactions of the least expert participants with DCR-HR and caused them to overlook some of the features of the DCR Portal. To mitigate this effect, all participants were taught the semantics of the DCR notation and were uniformly familiarized with the main features of the DCR Graphs Portal. Furthermore, participants with different levels of expertise might have exhibited slightly different behaviors compared to other participants in the same group. However, these differences did not clearly emerge during the analysis. The interactions between the researchers and participants is another possible threat to validity. To avoid this risk, a data collection protocol specifying all the steps of the data collection procedure has been followed during all sessions guaranteeing that all participants receive the same instructions and ensuring that the researcher is not biasing the insights provided by the participants.

External Validity. The design of the study, its exploratory nature, and the limited number of participants make the findings difficult to generalize. Nevertheless, the obtained insights allowed the identification of different reading patterns and strategies, which could be used as a basis for future investigations. In addition, the covered task types represent only a small subset of all the possible circumstances where process models are used in the real-world. Although the choice of these tasks was motivated by the input of experts in the field of process modeling, with close ties to legal practitioners and municipal employees, their elicitation was driven by practical motivations rather than theoretical foundations, which could be considered in future studies.

We could report two main strategies describing the behavior of people engaging DCR-HR and observed that people tend to become more goal-directed while becoming more acquainted with different artifacts. Although we were able to categorize the behavior of people into goal-directed and exploratory, we do not exclude that there are other strategies followed by people when engaging with hybrid process artifacts, such as those outlined in [39] considering a mix of goal-directed and exploratory behaviors depending on when and how comprehension tasks are shown to participants.

The hybrid process artifact we investigate in this work is based on DCR Graphs. Hence, the validity of our insights is bound to that specific language. However, the constraint-based approach applied in DCR is shared with many other declarative languages (e.g., Declare [15]). Therefore, our findings could presumably apply to other languages in the declarative paradigm as IT specialists would probably, like in our study, struggle reading the law text and domain experts would have a preference for the law text. It is however unclear whether our results can be generalized to imperative process models. An imperative model (e.g., in BPMN [47]), would have been possibly easier to understand for domain experts, and thus more interactions with the process model could have occurred.

In our exploratory study the hybrid artifacts were overlapping in terms of the information that each artifact provided. Indeed the simulation and the DCR Graph were information equivalent and the model reflected largely the specifications in the law text. It is rather unlikely that our findings would generalize to hybrid representations with limited information overlap between artifacts (e.g., imperative models enriched with business rules [50]). In this case it might be difficult for domain experts or IT specialists to rely on a single artifact to solve a particular task, but it might become necessary to understand all the artifacts. In such settings it becomes even more important to focus on the quality of the artifacts composing a hybrid representation, since it needs to be ensured that both domain experts and IT specialists can make sense of them.

Construct Validity. The measures used to answers our research questions (cf. Table 1) show differences between groups of participants when dealing with different tasks. The interpretation of these differences can cause a threat to validity if not correctly triangulated with other data sources and supported by existing literature. To reduce this threat, the interpretation of the eye-tracking measures was supported by the participants' verbal utterances and the notes collected from their eye gaze recordings before being linked to the existing body of knowledge in literature.

Reliability. Reproducibility is a crucial requirement for any empirical research. Although our findings are qualitative and use the subjective insights obtained from the participants, we have followed a systematic approach based on concepts from grounded theory (cf. Section 4.3) in order to identify the pertinent aspects evoked by the participants. Nevertheless, the coding procedure might entail some subjectivity which could have biased our findings. To reduce this effect, our coding was constantly reviewed and discussed by the co-authors, ensuring a consensus in coding borderline cases.

7 Conclusion and Future Work

This study takes an initial stride towards providing an in-depth understanding of the way declarative process models are read when combined with other artifacts. Looking at the way participants with different backgrounds engaged with DCR-HR during the execution of different tasks, we observed that municipal employees and academics used artifacts in a different way. Accordingly, the background of the users seems to influence the perceived benefits and challenges associated to each artifact, thus influencing the way they are used. Exploring how different artifacts are used over time, we noticed that people follow different strategies when engaging with DCR-HR and tend switch from an exploratory to a goal-directed behavior as time goes by.

The findings delineate clear directions for future work. First and foremost, it is necessary to investigate the different reading patterns and strategies identified throughout this study in light of the support they provide for a better understanding of the business process. To this end, performance metrics, such as answer accuracy and response time, could be correlated with the identified patterns to discern the most efficient ones. With the availability of more data, the association between reading patterns and understandability and its operationalization through performance metrics could be used in practice to develop a statistical model that could be trained to predict the performance of users based on the patterns they exhibit during a comprehension task.

Another relevant direction for future work is the exploitation of the identified reading patterns and strategies to provide better tool-support for users at run-time. Indeed, the behavioral features of the participants could be used at run-time to determine their background and expertise, and accordingly adjust the hybrid representation by highlighting the relevant areas on the artifacts they are most likely to use. This feature is expected to reduce the cognitive distraction, and the attention split [82] caused by the display of different artifacts. The identification of the task at hand could further reduce such a cognitive effect by putting more emphasis on the artifacts relevant for solving a particular task and shading the irrelevant ones. Automating the identification of the users' strategies at run-time could also contribute to an increased tool support. Herein, additional guidance could be provided to users with exploratory behavior by providing cues allowing to reduce the search area and thus promoting a rather goal-directed behavior.

Understanding the user behavior during a comprehension task can be extended beyond the proxy tasks covered by this study. Such an extension could be done by adopting the proposed qualitative approach in the analysis of the modeling and the maintaining of process models with the support of hybrid process artifacts. Herein, more robust analysis techniques supported by eye-tracking and other biosensor devices could be used to obtain fine-grained insights about the support of hybrid process artifacts, not only with regards to comprehension tasks but also when modeling and maintaining processes.

Besides, our exploratory insights might improve the design and the modeling of hybrid process artifacts. As the background of users influence their preferences for artifacts, it is important to consider the audience targeted by the hybrid representation and pinpoint the less clear aspects in the model requiring to be enriched with additional artifacts. In this vein, future work could investigate the ways in which process artifacts can overlap and propose clear guidelines on designing hybrid process artifacts. Moreover, our findings show that users tend to switch from an exploratory behavior to a goal-directed behavior progressively. Related to that, the quality of process artifacts could be further investigated, ensuring that novices can rapidly make sense of them and thus facilitating the transition from an exploratory behavior to a goal-directed behavior. This is particularly the case for the DCR Graph, which could be modeled in different ways. While many guidelines have emerged to prescribe the factors affecting the quality of imperative process models, little is known about the quality of declarative models. Future work could take this direction to investigate and infer the quality factors affecting the understandability of declarative models, particularly those represented in DCR Graphs.

Overall, the insights arising from this exploratory study are expected to have an important impact on current research. Indeed, the different reading patterns raise several questions about their influence on the comprehension of hybrid process artifacts and the underlying human cognitive processes. This paves the path for future investigations aimed to improve the design of these artifacts. Additionally, steering the direction for future work towards the development of adaptive tool-support, learning and adjusting to the users' behavior, would certainly help to bring the current research to practice by providing run-time support to users.

References

1. R Hevner Von Alan, Salvatore T March, Jinsoo Park, and Sudha Ram. Design science in information systems research. *MIS quarterly*, 28(1):75–105, 2004.
2. Frank Dengler and Denny Vrandečić. Wiki-Based Maturing of Process Descriptions. pages 313–328. 2011.
3. Jakob Pinggera, Thomas Porcham, Stefan Zugal, and Barbara Weber. LiProMo-Literate process modeling. *CEUR Workshop Proceedings*, 855:163–170, 2012.
4. Stefan Zugal, Jakob Pinggera, and Barbara Weber. Creating Declarative Process Models Using Test Driven Modeling Suite. In Selmin Nurcan, editor, *IS Olympics: Information Systems in a Diverse World*, pages 16–32. Springer Berlin Heidelberg, 2012.
5. Stefan Zugal, Jakob Pinggera, and Barbara Weber. Toward enhanced life-cycle support for declarative processes. *Journal of Software: Evolution and Process*, 24(3):285–302, apr 2012.
6. Johannes De Smedt, Jochen De Weerdt, Estefanía Serral, and Jan Vanthienen. Improving Understandability of Declarative Process Models by Revealing Hidden Dependencies. In Selmin Nurcan, Pnina Soffer, Marko Bajec, and Johann Eder, editors, *Advanced Information Systems Engineering*, pages 83–98, Cham, 2016. Springer International Publishing.
7. Johannes De Smedt, Jochen De Weerdt, Estefanía Serral, and Jan Vanthienen. Discovering hidden dependencies in constraint-based declarative process models for improving understandability. *Information Systems*, 74:40–52, may 2018.
8. Hugo A López, Søren Debois, Thomas T Hildebrandt, and Morten Marquard. The process highlighter: From texts to declarative processes and back. In *BPM (Dissertation/Demos/Industry)*, pages 66–70, 2018.
9. Amine Abbad Andaloussi, Jon Buch-Lorentsen, Hugo A. López, Tijs Slaats, and Barbara Weber. Exploring the modeling of declarative processes using a hybrid approach. In Alberto H. F. Laender, Barbara Pernici, Ee-Peng Lim, and José Palazzo M. de Oliveira, editors, *Conceptual Modeling*, pages 162–170, Cham, 2019. Springer International Publishing.
10. Amine Abbad Andaloussi, Andrea Burattin, Tijs Slaats, Anette Chelina Møller Petersen, Thomas T. Hildebrandt, and Barbara Weber. Exploring the understandability of a hybrid process design artifact based on dcr graphs. In Iris Reinhartz-Berger, Jelena Zdravkovic, Jens Gulden, and Rainer Schmidt, editors, *Enterprise, Business-Process and Information Systems Modeling*, pages 69–84, Cham, 2019. Springer International Publishing.
11. Amine Abbad Andaloussi, Andrea Burattin, Tijs Slaats, Ekkart Kindler, and Barbara Weber. On the declarative paradigm in hybrid business process representations: A

conceptual framework and a systematic literature study. *Information Systems*, 91:101505, 2020.
12. Manfred Reichert and Barbara Weber. *Enabling Flexibility in Process-Aware Information Systems.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
13. Stefan Zugal, Jakob Pinggera, and Barbara Weber. The impact of testcases on the maintainability of declarative process models. In Terry Halpin, Selmin Nurcan, John Krogstie, Pnina Soffer, Erik Proper, Rainer Schmidt, and Ilia Bider, editors, *Enterprise, Business-Process and Information Systems Modeling*, pages 163–177, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
14. Johannes De Smedt, Jochen De Weerdt, Jan Vanthienen, and Geert Poels. Mixed-paradigm process modeling with intertwined state spaces. *Business and Information Systems Engineering*, 58(1):19–29, 2016.
15. Maja Pesic, Helen Schonenberg, and Wil M.P. van der Aalst. DECLARE: Full Support for Loosely-Structured Processes. In *11th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2007)*, pages 287–287. IEEE, oct 2007.
16. Søren Debois, Thomas Hildebrandt, Morten Marquard, and Tijs Slaats. Hybrid Process Technologies in the Financial Sector. *Proceedings of the Industry Track at the 13th International Conference on Business Process Management 2015*, 4:107–119, 2015.
17. Thomas T. Hildebrandt and Raghava Rao Mukkamala. Declarative Event-Based Workflow as Distributed Dynamic Condition Response Graphs. *Electronic Proceedings in Theoretical Computer Science*, 69:59–73, 2011.
18. Morten Marquard, Muhammad Shahzad, and Tijs Slaats. Web-Based Modelling and Collaborative Simulation of Declarative Processes. In Hamid Reza Motahari-Nezhad, Jan Recker, and Matthias Weidlich, editors, *Business Process Management*, pages 209–225, Cham, 2016. Springer International Publishing.
19. K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. *Eye Tracking: A comprehensive guide to methods and measures.* OUP Oxford, 2011.
20. Jennifer Romano Bergstrom and Andreei Schall. *Eye tracking in user experience design.* Elsevier, 2014.
21. Georg Buscher, Andreas Dengel, Ralf Biedert, and Ludger V Elst. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(2):9, 2012.
22. Wil M.P. Van Der Aalst. *Process mining: Discovery, Conformance, and Enhancement of Business processes.* Springer International Publishing, 2010.
23. Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. Visualization of eye tracking data: A taxonomy and survey. In *Computer Graphics Forum*, volume 36, pages 260–284. Wiley Online Library, 2017.
24. Chris Janiszewski. The Influence of Display Characteristics on Visual Exploratory Search Behavior. *Journal of Consumer Research*, 25(3):290–301, 12 1998.
25. Jella Pfeiffer, Martin Meißner, Jascha Prosiegel, and Thies Pfeiffer. Classification of goal-directed search and exploratory search using mobile eye-tracking. *Proceedings of the International Conference on Information Systems*, 2014.
26. W. Maalej, M. Nayebi, T. Johann, and G. Ruhe. Toward data-driven requirements engineering. *IEEE Software*, 33(1):48–54, Jan 2016.
27. Tijs Slaats. Declarative and hybrid process discovery: Recent advances and open challenges. *Journal on Data Semantics*, pages 1–18, 2020.
28. M. Pesic. *Constraint-based workflow management systems : shifting control to users.* PhD thesis, TU Eindhoven, 2008.
29. Dirk Fahland, Daniel Lübke, Jan Mendling, Hajo Reijers, Barbara Weber, Matthias Weidlich, and Stefan Zugal. Declarative versus Imperative Process Modeling Languages: The Issue of Understandability. In *Proceedings of EMMSAD*, pages 353–366. Springer, 2009.
30. Dirk Fahland, Jan Mendling, Hajo A Reijers, Barbara Weber, Matthias Weidlich, and Stefan Zugal. Declarative versus Imperative Process Modeling Languages: The Issue of Maintainability. In Stefanie Rinderle-Ma, Shazia Sadiq, and Frank Leymann, editors, *Business Process Management Workshops*, pages 477–488, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
31. George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.
32. Søren Debois, Thomas Hildebrandt, Morten Marquard, and Tijs Slaats. Bridging the valley of death: A success story on danish funding schemes paving a path from technology readiness level 1 to 9. In *Proceedings of the Second International Workshop on Software Engineering Research and Industrial Practice*, pages 54–57, Piscataway, NJ, USA, 2015. IEEE Press.
33. Tijs Slaats. *Flexible Process Notations for Cross-organizational Case Management Systems.* PhD thesis, IT University of Copenhagen, January 2015.
34. Søren Debois, Thomas T Hildebrandt, Morten Marquard, and Tijs Slaats. The dcr graphs process portal. In *BPM (Demos)*, pages 7–11, 2016.
35. Kuno Kurzhals, Brian Fisher, Michael Burch, and Daniel Weiskopf. Eye tracking evaluation of visual analytics. *Information Visualization*, 15(4):340–358, 2016.
36. Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*, pages 45–52. ACM, 2012.
37. Jens Gulden, Andrea Burattin, Amine A. Andaloussi, and Barbara Weber. From analytical purposes to data visualizations: a decision process guided by a conceptual framework and eye tracking. *Software & Systems Modeling*, Jul 2019.
38. Kuno Kurzhals, Brian Fisher, Michael Burch, and Daniel Weiskopf. Evaluating visual analytics with eye tracking. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 61–69, New York, NY, USA, 2014. ACM.
39. Fang Zhao, Wolfgang Schnotz, Inga Wagner, and Robert Gaschler. Eye tracking indicators of reading approaches in text-picture comprehension. *Frontline Learning Research*, 2(5):46–66, 2014.
40. Stefan Zugal. *Applying Cognitive Psychology for Improving the Creation, Understanding and Maintenance of Business Process Models.* PhD thesis, University of Innsbruck, 2013.
41. Kathrin Figl. Comprehension of procedural visual business process models. *Business & Information Systems Engineering*, 59(1):41–67, Feb 2017.
42. F. Garcia, F. Ruiz, and C. A. Visaggio. A proposal and empirical validation of metrics to evaluate the maintainability of software process models. In *2006 IEEE Instrumentation and Measurement Technology Conference Proceedings*, pages 1093–1097, April 2006.

43. Jan Claes, Irene Vanderfeesten, Frederik Gailly, Paul Grefen, and Geert Poels. The structured process modeling theory (spmt) a cognitive view on why and how modelers benefit from structuring the process of process modeling. *Information Systems Frontiers*, 17(6):1401–1425, Dec 2015.
44. Stefan Zugal, Cornelia Haisjackl, Jakob Pinggera, and Barbara Weber. Empirical evaluation of test driven modeling. *International Journal of Information System Modeling and Design*, 4(2):23–43, April 2013.
45. Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, Berlin, Heidelberg, 2007.
46. Razvan Petrusel and Jan Mendling. Eye-tracking the factors of process model comprehension tasks. In Camille Salinesi, Moira C. Norrie, and Óscar Pastor, editors, *Advanced Information Systems Engineering*, pages 224–239, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
47. Object Management Group OMG. Business Process Modeling Notation V 2.0. 2006.
48. Palash Bera, Pnina Soffer, and Jeffrey Parsons. Using eye tracking to expose cognitive processes in understanding conceptual models. *MIS Quarterly: Management Information Systems*, 43:1105–1126, 2019.
49. G Keller, M Nüttgens, and A W Scheer. *Semantische Prozessmodellierung auf der Grundlage "ereignisgesteuerter Prozessketten (EPK)"*. Institut für Wirtschaftsinformatik Saarbrücken: Veröffentlichungen des Instituts für Wirtschaftsinformatik. Inst. für Wirtschaftsinformatik, 1992.
50. Wei Wang, Marta Indulska, Shazia Sadiq, and Barbara Weber. Effect of linked rules on business process model understanding. In *Business Process Management*, pages 200–215. Springer International Publishing, Cham, 2017.
51. B. Sapkota and M. v. Sinderen. Exploiting rules and processes for increasing flexibility in service composition. In *2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops*, pages 177–185, Oct 2010.
52. Michael Zimoch, Rüdiger Pryss, Johannes Schobel, and Manfred Reichert. Eye tracking experiments on process model comprehension: Lessons learned. In Iris Reinhartz-Berger, Jens Gulden, Selmin Nurcan, Wided Guédria, and Palash Bera, editors, *Enterprise, Business-Process and Information Systems Modeling*, pages 153–168, Cham, 2017. Springer International Publishing.
53. Cornelia Haisjackl, Irene Barba, Stefan Zugal, Pnina Soffer, Irit Hadar, Manfred Reichert, Jakob Pinggera, and Barbara Weber. Understanding declare models: strategies, pitfalls, empirical results. *Software and Systems Modeling*, 15(2):325–352, 2016.
54. Cornelia Haisjackl, Pnina Soffer, Shao Yi Lim, and Barbara Weber. How do humans inspect BPMN models: an exploratory study. *Software and Systems Modeling*, 17(2):655–673, 2018.
55. Jakob Pinggera. *The Process of Process Modeling*. PhD thesis, Department of Computer Science, 2014.
56. Jakob Pinggera, Pnina Soffer, Dirk Fahland, Matthias Weidlich, Stefan Zugal, Barbara Weber, Hajo A Reijers, and Jan Mendling. Styles in business process modeling: an exploration and a model. *Software & Systems Modeling*, 14(3):1055–1080, 2015.
57. Victor R Basili Gianluigi Caldiera and H Dieter Rombach. The goal question metric approach. *Encyclopedia of software engineering*, pages 528–532, 1994.
58. K. Charmaz. *Constructing Grounded Theory*. Introducing Qualitative Methods series. SAGE Publications, 2014.
59. Daniel C Richardson and Rick Dale. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6):1045–1060, 2005.
60. A Olsen. The tobii i-vt fixation filter. *Tobii Technology*, 2012.
61. IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams. *IEEE Std 1849-2016*, pages 1–50, Nov 2016.
62. A. J. M. M. Weijters, Wil M P van der Aalst, and A. K. Alves De Medeiros. Process Mining with the Heuristics Miner Algorithm. *TU/e Tech. Report*, 166:1–34, 2006.
63. Kevin R. Coombes, Guy Brock, and Zachary B. Abrams. Polychrome: Creating and assessing qualitative palettes with many colors. *Journal of Statistical Software*, 2019.
64. Jens Gulden. Visually comparing process dynamics with rhythm-eye views. In *BPM Workshops*, LNBIP, pages 474–485. Springer, 2016.
65. J. Saldana. *The Coding Manual for Qualitative Researchers*. SAGE Publications, 2012.
66. Susanne Friese. *Qualitative data analysis with ATLAS. ti*. SAGE Publications Limited, 2019.
67. Claudia Ehmke and Stephanie Wilson. Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*, pages 119–128. British Computer Society, 2007.
68. Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Fundamentals of Business Process Management*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2018.
69. Mackenzie G. Glaholt, Mei-Chun Wu, and Eyal M. Reingold. Evidence for top-down control of eye movements during visual decision making. *Journal of Vision*, 10(5):15–15, 05 2010.
70. Andrea Burattin, Michael Kaiser, Manuel Neurauter, and Barbara Weber. Learning process modeling phases from modeling interactions and eye tracking data. *Data and Knowledge Engineering*, 121:1 – 17, 2019.
71. Avner Ottensooser, Alan Fekete, Hajo A. Reijers, Jan Mendling, and Con Menictas. Making sense of business process descriptions: An experimental comparison of graphical and textual notations. *Journal of Systems and Software*, 85(3):596 – 606, 2012. Novel approaches in the design and implementation of systems/software architecture.
72. B. Curtis, S. B. Sheppard, E. Kruesi-Bailey, J. Bailey, and D. A. Boehm-Davis. Experimental evaluation of software documentation formats. *J. Syst. Softw.*, 9(2):167–207, February 1989.
73. Wolff-Michael Roth and Gervase Michael Bowen. When are graphs worth ten thousand words? an expert-expert study. *Cognition and Instruction*, 21(4):429–473, 2003.
74. Thomas G. Moher, Dr. K. H. Mak, B. B. Blumenthal, and L. M. Levanthal. Comparing the comprehensibility of textual and graphical programs. 1993.
75. Cornelia Haisjackl and Stefan Zugal. Investigating differences between graphical and textual declarative process models. In Lazaros Iliadis, Michael Papazoglou, and Klaus Pohl, editors, *Advanced Information Systems Engineering Workshops*, pages 194–206, Cham, 2014. Springer International Publishing.
76. Jill H. Larkin and Herbert A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1):65 – 100, 1987.

77. Daniel Moody. The “physics” of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Trans. Softw. Eng.*, 35(6):756–779, November 2009.
78. Keng Siau. Informational and computational equivalence in comparing information modeling methods. *J. Database Manag.*, 15:73–86, 01 2004.
79. Allan Paivio. Dual coding theory: Retrospect and current status. 1991.
80. Certainly We Were Not Many and Yet We Were Sufficient. Estimating the number of subjects needed for a thinking aloud test. *Int. J. Human-Computer Studies*, 41:385–397, 1994.
81. Gay Costain. *Cognitive support during object-oriented software development: the case of UML diagrams.* PhD thesis, ResearchSpace@ Auckland, 2008.
82. Slava Kalyuga, Paul Ayres, Paul Chandler, and John Sweller. The expertise reversal effect. *Educational Psychologist*, 38(1):23–31, 2003.