Comprehending 3D and 4D ontology-driven conceptual models: An empirical study

Verdonck, M., Gailly, F. and De Cesare, S.

Comprehending 3D and 4D ontology-driven conceptual models: an empirical study

Michaël Verdonck1, Frederik Gailly1 and Sergio de Cesare2

1 Faculty of Economics and Business Administration, Ghent University, Belgium
2 University of Westminster, London, UK

michael.verdonck@ugent.be; frederik.gailly@ugent.be; s.decesare@westminster.ac.uk

Abstract: This paper presents an empirical study that investigates the extent to which the pragmatic quality of ontology-driven models is influenced by the choice of a particular ontology, given a certain understanding of that ontology. To this end, we analyzed previous research efforts and distilled three hypotheses based on different metaphysical characteristics. An experiment based on two foundational ontologies (UFO and BORO) involving 158 participants was then carried out, followed by a protocol analysis to gain further insights into the results of experiment. We then extracted five derivations from the results of the empirical study in order to summarize our findings. Overall, the results confirm that the choice of a foundational ontology can lead to significant differences in the interpretation and comprehension of the conceptual models produced. Moreover, the effect of applying a certain foundational ontology can cause considerable variations in the effort required to comprehend these models.

1 Introduction

As defined by Mylopoulos [1], conceptual modeling is the activity of representing aspects of the physical and social world for the purpose of communication, learning and problem solving among human users. Over the years, conceptual modeling has become a fundamental discipline in research fields such as Knowledge Engineering, Knowledge Representation and Information Modeling. Considering that poor system design and planning is cited as one of the most determining factors in systems failures [2], the importance of design has been well recognized in the information systems domain. Especially in enterprises, conceptual modeling has been applied frequently for the development of information systems, business processes and enterprise architecture [3, 4]. Since a conceptual model is considered as a key means for communication, analysis and documentation of domain knowledge and system requirements, the quality of the resulting conceptual model can have a considerable influence on the quality of the developed system or process [5]. Criticism however arose, stating that conceptual models lacked a foundational theory to define a domain and its concepts, which led to duplication of information, diffuse interpretations and contradictory requirements [6]. As a result, research proposed a new approach named **ontology-driven conceptual modeling** (**ODCM**). Through the use of ontologies, ODCM allows to create semantically stronger conceptual models, leading to better representations of the systems to be, which can be used in order to build more reliable information systems. Ontological theories support the construction of explicit models of reality, by offering guidelines and instructions to select the appropriate concepts which should be represented as language constructs [7]. More specifically, in this paper we shall refer to all techniques where ontologies are applied – e.g. evaluation, analysis or theoretical foundation – to improve either the quality of the conceptual modeling process or the quality of the conceptual model, as ODCM. Based upon their level of dependence of a particular task or point of view, different types of ontologies can be distinguished and applied in ODCM [8]. In this article however, we will focus solely on foundational ontologies – which can be defined as ontologies that define a range of top-level domain-independent ontological categories, which form a general foundation for more elaborated domain-specific ontologies [9].

Different kinds of foundational ontologies can be adopted in order to perform ODCM. For instance, based upon the endurantism-perdurantism paradigms, we can differentiate between 3D and 4D ontologies. *3D ontologies* view individual objects as three-dimensional, having only spatial parts, and wholly exist at each moment of their existence. *4D ontologies* idiosyncratically see individual objects as four-dimensional, having spatial and temporal parts, and exist immutably in space-time [10]. While most research in ODCM has been performed with 3D ontologies [11], 4D ontologies have gained more popularity in recent years [12–14]. Despite the research of de Cesare et al. [13] and Hadar and Soffer [15] whose efforts have already demonstrated that applying different ontologies can lead to diverse kinds of conceptualizations, there exists little research that profoundly investigates the impact of applying these different kinds of ontologies on the resulting conceptual models. Furthermore, while ontologies were introduced to increase the overall quality of conceptual models, past research has mainly emphasized the semantic quality of models, and has spent little effort in examining the comprehension of models [16, 17]. As a result, there is a high academic interest in further exploring and discovering the differences between adopting different kinds of foundational ontologies, as demonstrated for instance by the 6th and 7th International Workshop on Ontologies and Conceptual Modeling (OntoCom)[1,2] which dedicated the full program on this particular topic at the 38th International Conference on Conceptual Modeling and the 10th International Conference on Formal Ontology in Information Systems.

Therefore, to further examine this topic, this paper will perform a rigorous investigation of the effects of applying different kinds of foundational ontologies on the comprehension of their resulting models – also known as pragmatic quality [18]. To properly measure these effects, we conduct an empirical study. As the foundation for the further development of this paper, we formulate our research question as follows: *Is the pragmatic quality of an ontology-driven conceptual model influenced by the choice of a particular foundational ontology, given a certain understanding of that foundational ontology?* In other words, we investigate the influence of ontology on the interpretation and understanding of the resulting conceptual models by taking into account the pre-existing knowledge a person has of the respective ontology. In section 2 of this paper, we will discuss in more depth the research related to ontology-driven conceptual modeling and 3D and 4D ontologies. Section 3 will explain the research methodology and design of our empirical study, where we also formulate our hypotheses. Next, we draft our experimental design to test these hypotheses in section 4. We then present the results of our experiment in section 5 and discuss the acceptance or rejection of our hypotheses. To better understand these results, we perform a protocol analysis, of which the design and the results will be discussed in section 6. Next, in section 7, we will discuss the consequences and implications of the results from both the experiment and the protocol analysis and provide an answer to our research question. Finally, we will present our conclusion and future research opportunities in section 8.

[1] OntoCom 2018: http://ceur-ws.org/Vol-2205/
[2] OntoCom 2019: https://www.springer.com/gp/book/9783030341459

2 Related Research

Ontology-driven conceptual modeling

While 'traditional' conceptual modeling can be described as the activity of representing aspects of the physical and social world for the purpose of communication, learning and problem solving among human users [1], ODCM can be characterized as the utilization of ontological theories, coming from areas such as formal ontology, cognitive science and philosophical logics, to develop engineering artifacts (e.g. modeling languages, methodologies, design patterns and simulators) for improving the theory and practice of conceptual modeling [19]. The benefits of ODCM are considered to be the most profound when applied for the design, analysis and re-engineering of rather large and complex information systems. Their use would lead to various system engineering benefits such as increased reusability and reliability [20]. Moreover, ODCM can aid in the development of a more sophisticated representation of the modeling of a domain, and result in a higher level of domain understanding by its modelers and users [21]. A wide variety of techniques and practices can be related to ODCM. For example, ontologies were used for the development of new conceptual modeling languages [22], for adding structuring rules to existing languages [23], and for proposing conceptual modeling patterns and anti-patterns [24]. Several techniques in ODCM go further than only evaluating or supporting a conceptual modeling technique, but also provide a conceptual modeling technique themselves, as such adapting modelers to an ontological way of thinking. Such techniques (e.g. OntoUML, O3) are often founded on existing modeling notations and enhance the metamodel of this notation by incorporating formal ontological constraints that correspond to the ontology's axiomatization. It has been observed [25] that (novice) modelers applying such ODCM techniques arrive at higher quality models compared to modelers applying more traditional conceptual modeling techniques. By applying these ontology-driven techniques, modelers are encouraged to reason more carefully about the domain and the classification of its elements in terms of ontological concepts (e.g. inferring types, properties, events etc.) and their underlying relationships. The above research efforts demonstrate the benefits of adopting

ODCM. Nonetheless, while these benefits are clearly intertwined with the choice of ontological theory that serves as a foundation for the respective ODCM technique, there exists little research that investigates the impact of applying different kinds of ontologies when adopting ODCM. Based upon their level of dependence of a particular task or point of view, different types of ontologies can be distinguished and applied in ODCM such as domain ontologies, core ontologies and foundational ontologies [8]. In this article however, we will focus solely on foundational ontologies since these describe very general concepts – like space, time, matter, object, event, action, etc. – and are independent of a particular problem or domain. Moreover, we will focus on two well-known paradigms that exist within foundational ontologies: endurantist and perdurantist ontologies or also known respectively as 3D and 4D ontologies.

3D and 4D ontologies

In endurantism or 3D ontology, an individual thing such as for example John Doe endures through time and is regarded as totally present at any moment in its lifetime. In a perdurantist or 4D ontology, an individual thing perdures through time and is extended in time, and so can be said to be only partially present at any moment in time – e.g. the whole of John extends over time from his birth to his death. Thus, while 3D ontologies view objects only from the present and assume that the same object can exist over time and may be fully identified at different points in time, the 4D ontological view emphasizes the continuity of objects over space-time, where these objects exist immutably. These differences between endurants and perdurants determine whether and how objects exist in different ways in the past, present and future. Or in other words these differences determine how an ontology is formed.

While most research efforts in ODCM can be found with 3D ontologies as the underlying foundation [11], 4D ontologies have become more adopted in recent years [12–14]. The effects of adopting a 3D or a 4D ontology with an ODCM technique and their impact on the resulting conceptual models however are still unclear and little research has yet profoundly investigated their differences. The research efforts that have been performed do demonstrate that applying different ontologies can lead to diverse kinds of conceptualizations. For instance, the research of de Cesare et al. [13] examined the differences between 3D and 4D ontologies and their quality in representing temporal changes, with a specific focus on roles. They found that two key differences, i.e. the different criteria of identity and the differences in the way individual objects extend over time and possible worlds, have an affect on the way roles are being modelled between both ontologies, concluding that the 4D approach is best able to describe temporal changes wherein roles are spatio-temporal extents of individuals. Another study performed by Hadar and Soffer [15], analyzed two modeling frameworks each based on a 3D and 4D ontology respectively and evaluated their potential contribution to a reduction in conceptual modelling variations and thus their ability to facilitate model understanding. Their findings highlight contradictions in the guidance provided by the different frameworks, where differences in the underlying foundational ontologies exist. Hence, as these prior research efforts have demonstrated, it appears that a model will differ depending on the ontology that has been applied. However, since little research has been performed in this area, limited knowledge exists on the fundamental differences between applying different ontologies on such models.

Therefore, we will perform an empirical comparison, more specifically between a 3D and a 4D foundational ontology. Our choice for these two kinds of ontologies is two-fold. First, although there exists much theoretical work on both types of ontologies, there has not yet been an empirical comparison between them. While both have their advantages and disadvantages, and consequently influence the conceptualizations that are realized by these ontologies, there has not yet been a study to test if these conceptualizations actually lead to significant differences in the pragmatic quality of their resulting conceptualizations. Second, to perform a comparison between ontologies, it is desirable that these ontologies be considerably different from one another. Since 3D and 4D ontologies originate from quite different paradigms, this will result in different kinds of models. In order to clearly distinguish between these ontologies, we will focus on their metaphysical characteristics, and the influence of these particular characteristics on the comprehension of the resulting ontology-driven models.

The UFO and BORO ontologies

In this study, we decided to compare the Unified Foundational Ontology (UFO) (a 3D ontology) and the Business Object Reference Ontology (BORO) (a 4D ontology). Our choice for these two specific ontologies is driven by various reasons. First, they are both foundational ontologies that are repeatedly applied in ODCM. Secondly, both ontologies can be further differentiated based on their difference in utilization and their purpose, making them more interesting to compare. For example, UFO was developed for analyzing modeling languages and to improve them. More specifically, the aim of UFO is to improve the truthfulness to reality (domain appropriateness) and conceptual clarity (comprehensibility appropriateness) of a modeling language [26]. BORO, on the other hand, was developed for re-engineering purposes and to integrate systems in a transparent and straightforward manner [27]. By utilizing business objects, its purpose is to make systems simpler and functionally richer so that in practice, they would be cheaper to build and maintain. While we will not cover all the concepts of both ontologies in this paper, we do encourage the interested reader for a more detailed description of these ontologies for BORO to [27, 28] and for UFO to [19, 29]. A brief description of both ontologies however can also be found in the online repository that accompanies this paper at Open Science Framework (OSF)[3].

BORO distinguishes three main categories [30]: Individuals, Types, and tuples. Every object belongs to one of these categories. Elements are individual objects whose identity is given by the element's spatiotemporal extent i.e. the space and time it occupies. An example of an element would be the person John. Types are collections of any type of object (in other words, objects of any of the three categories). The identity of a type is determined by its extension, the collection of its instances or members. For example, the extension of the type Persons is the set of all people. Individuals can be collected into types, referred to as IndividualTypes. Finally, Tuples are relations between objects. The identity of a tuple is defined by the places in the tuple. An example is (Mary, John) in which the elements Mary and John occupy places 1 and 2 in the

[3] https://osf.io/ahfjc/

tuple, respectively. Similar to Individuals, Tuples can be collected into types, called TupleTypes. An example is parentOf, which is the collection of all relations between parents and their children.

UFO is a much more complex ontology, being composed of three core modules, namely UFO-A, an ontology of endurants (objects) [29]; UFO-B, an ontology of perdurants (events) [31]; and UFO-C and ontology of social entities that specializes the former two [32]. UFO's most fundamental distinction is between individuals and universals. Universals are abstract patterns of features that can be realized in a number of different individuals. In UFO, the identity of universals is separated from their extensions, allowing, for instance, that the extension of a universal changes through time, as well as that two universals have the same extension at a given moment. A second fundamental distinction in UFO is that between endurants and perdurants. Endurants are individuals that are wholly present whenever they are present (e.g a ball, a person), whilst perdurants are individuals extended in time, through which they accumulate temporal parts (e.g. a football match, a party). UFO further categorises endurants into substantials and moments. Substantials are existentially-independent endurants (e.g., an animal, a table), whilst moments are individuals that necessarily inhere in others, their bearers, to which they are existentially-dependent on (e.g., the color of an object, the weight of a person). In UFO, endurants can be characterized by both essential and contingent properties, which means that they can genuinely change while keeping their identity [29]. For instance, being a mammal is an essential property for dogs, whilst being overweight is a contingent one. Thus, no dog can change in such a way that it ceases to be a mammal, whilst any dog can change its weight and remain the same individual. Perdurants, conversely, are exclusively characterized by essential properties, thus they cannot genuinely change) [31]. In the section below, we will discuss the different metaphysical choices that shape the fundamentals of an ontology and derive our hypotheses from these metaphysical characteristics.

3 Methodology and Hypotheses Development

3.1 Methodology overview

This empirical research is part of a project that has been in development for several years. In a first research effort [33], we theoretically examined the principal conceptual model variations that resulted from constructing different enterprise models with a 3D and a 4D ontology. Since the resulting representations differed quite substantially from one another, we decided a further investigation into these differences in the form of an exploratory analysis [34]. More specifically, the exploratory study focused on the comprehension and understandability of ontology-driven models that were developed with either a 3D or 4D foundational ontology. Similarly to Mendling et al. [35], we regarded comprehension as the understanding of users on what the grammatical elements in a conceptual model mean and the actual understanding of the reality that the overall conceptual model represents. The results from this exploratory analysis confirmed that the conceptualizations that were realized by the different ontologies have a considerable impact on the understanding and comprehension of its users. Furthermore, the findings suggested that depending on the metaphysical characteristics of an ontology, some ontology-driven models are perceived as easier or more difficult to comprehend. However, we are still left with the question of the extent to which the pragmatic quality of these ontology-driven models is influenced by the choice of a particular ontology. Does the choice of an ontology have a significant effect on the interpretation and comprehension of the resulting models, or is this effect only marginal?

Hence, our previously conducted studies led to the formation of our research question as formulated in the introduction above. To address the research question, an empirical study is conducted based upon the experimental design described in Wohlin et al. [36]. The empirical study, as also depicted in Figure 1, is rendered as follows: first, we define our hypotheses that will serve as the basis for our empirical study. These hypotheses are based upon related research. Next, we will perform our experiment to test these hypotheses. The sole purpose of the experiment is to collect data to either accept or reject the hypotheses. Finally, in order to provide additional insights into the results of our experiment, we perform a protocol analysis. The purpose of the protocol analysis is to collect data which can help to interpret *why* the hypotheses were rejected or accepted. Thus, while the empirical study provides the data to accept or reject the hypotheses, the protocol analysis aims to provide additional insights into the nature of our results and can therefore be seen as a more in-depth analysis. This is similar to the research efforts of [37–39], where also their goal of the protocol analysis was not to produce data, but to acquire knowledge on how subjects perceived the experiment. A further distinction between both studies is that the empirical study is of a quantitative nature, while the protocol study is a qualitative method. More specifically, the empirical study focuses on measurement, and aims to isolate the pragmatic quality related to the ontology-driven conceptual models of both ontologies through dedicated instruments that collect scores in order to compare our two different treatments objectively. The protocol study on the other hand is a qualitative method designed to aid our understanding of the results of the empirical study in context. More specifically, the protocol study analyses verbal reports from participants subjects to reveal the mental processes taking place as individuals work on the interpretation of the models. More information on the specific experimental designs of each of these studies can be found in sections 4 and 6.

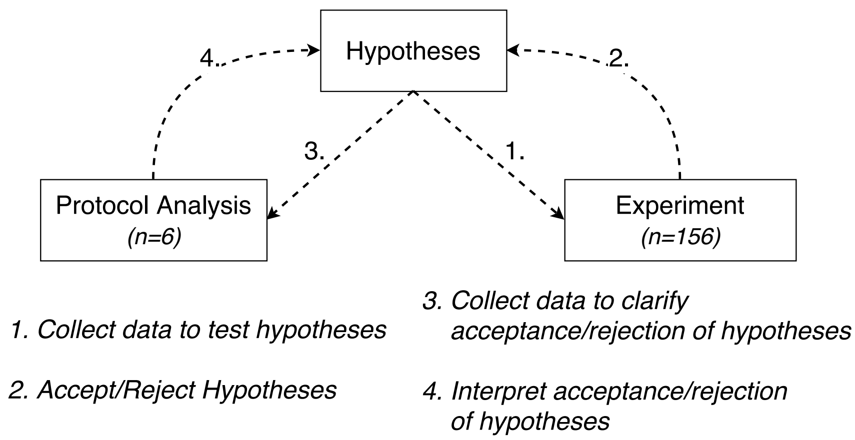

Figure 1: Method of performing the empirical study (n = number of subjects)

3.2 Hypotheses Development

We can distinguish between different ontologies based on their metaphysical characteristics. The metaphysical characteristics define the philosophical concepts and structures such as space, time, matter, object and event, and how these concepts interrelate with one another [40, 41]. Every ontology has their own metaphysical characteristics and represents real world phenomena in their specific way. Depending on these characteristics, an ontology can emphasize certain elements or structures such as time or identity, which can then influence the final representation of a conceptualization. The relevance of these influences has already been demonstrated in the domain of conceptual modeling. For example, in the research of Al Debei et al. [12], the 3D object-role modeling (ORM) paradigm was analytically compared to the 4D object paradigm (OP). The conducted comparison reveals that the OP paradigm can provide semantically richer representations of phenomena than the ORM paradigm. Also [13] and our initial research efforts [33, 42], theoretically examined the way in which a 3D ontology and a 4D ontology represent temporal changes, concluding that each of the ontologies can lead to different representations and interpretations. Regarding the main differences between 3D and 4D ontologies, each have their unique ontological interpretation of the following three metaphysical characteristics:

1. ***Notion of Identity***

The notion of identity and essence-defining characteristics – also referred to as essential properties in 3D ontologies – regards how the ontology assigns a principle of identity to its concepts and dictates whether a certain phenomenon is represented through one or multiple objects. In 4D ontologies, which follow a perdurantist view, individual objects are four-dimensional and have spatial and temporal parts. In a 3D ontology, the identity of an individual object is defined by its essential properties, or attributes that derive from a defining kind [43]. Unlike 3D ontologies, individual objects in a 4D ontology do not (explicitly) have properties such as color, weight or height. Instead, their features and characteristics are derived from relationships with other objects, for example, if a specific car is red then this car instantiates the type Red. Moreover, the difference concerning the notion of identity between 3D and 4D ontologies can be demonstrated through the representation of temporary conditions such as roles, states and phases of an element. For instance, in a 4D ontology, John as a child and John as an adult are separate individuals (i.e. states of John), that are related to John via a whole-part relationship. Conversely, in a 3D ontology, John as a child is the same as John as an adult since the essential properties of John do not change. In a 4D ontology, John is wholly present from his birth to his death. John as a child and John as an adult represent temporal parts of John as a whole. Hence, a 4D ontology has a straightforward general criterion of identity for all elements, i.e. their extension. A 3D ontology on the other hand relies on (multiple) criteria of identity, depending on its disjoint kinds. As a result, it has been argued that it can be difficult to articulate the identity criterion exactly in certain cases with a 3D ontology; a person being one such example. For instance, hair color, weight or length are obviously not essential properties of a person if we consider extended periods of time [44]. Thus, while the 3D ontological view has been argued to be more intuitive than the 4D ontological view in which objects exist immutably in space-time [44], it poses several difficulties concerning the principle of identity of 3D objects. Therefore, we define our first hypothesis as follows:

- ♦ *H_1: The notion of identity is more difficult to comprehend with 3D ontology-driven models than with 4D ontology-driven models, given a certain understanding of the respective ontology.*

2. *Perception and endurance of time*

The metaphysical choice of an ontology concerning time determines how entities begin and cease to exist over time, and how they define events and changes. In a 4D ontology, objects (including relationships) are represented immutably in space-time while 3D ontologies represent these objects and relationships in the present, with their current traits and characteristics. In 3D ontologies time-related elements are often represented through Events, which are individuals composed of temporal parts. They happen in time in the sense that they extend in time accumulating temporal parts. An example of 3D interactions between events and their relationships are described by the base relations in Allen's interval algebra for temporal reasoning [45]. With temporal reasoning, Events can be atomic or complex, depending on their mereological structure. Whilst atomic events have no proper parts, complex events are aggregations of at least two disjoint events. This composition implies that whenever an event is present, it is not necessarily the case that all its temporal parts are present. For instance, childhood and adulthood can be considered as Complex Events that are composed of a series of disjoint Atomic Events. As stated above, an element in a 4D ontology is defined by its spatiotemporal extent – i.e. the space and time it occupies – where the element extends through time and is not fully present at any instant in time (excluding elements with a zero-temporal extent). As such, the notion of time is an integral aspect in the definition of an individual object in a 4D ontology. Moreover, different segments of space-time are elements themselves. For instance, a temporal slice (e.g. a second, a day or a century) can be seen as a 4D extension. As stated by Hales & Johnson [29], 4D ontologies specifically emphasize the continuity of objects over space-time which they argue should result in 4D ontologies being more suitable to represent time-related concepts than 3D ontologies. Also de Cesare et al. [46] illustrated in their research how a 4D ontology is more appropriate to represent temporality and modality when modeling roles in comparison with a 3D ontology. Hence, we define our second hypothesis concerning the perception and endurance of time as follows:

- *H_2: The perception of time is more difficult to comprehend with 3D ontology-driven models than with 4D ontology-driven models, given a certain understanding of the respective ontology.*

3. ***The formation of relations***

This notion describes the nature of relationships between entities and how they form including parthood relationships between different entities. Similar to the notion of identity, the metaphysical choice of ontologies concerning the formation of relationships defines the way entities can become part of each other or separate from one another. In 3D ontologies, relationships can be distinguished based upon a certain meaning that is derived from the kinds of entities they link together. Relationships in 3D ontologies are in essence an objectification of a relational property and are thus existentially dependent on a multitude of different individuals. Foundational ontologies that typically adopt a 4D paradigm are extensionalist in the way that they define the identity of individual entities, but also the identity of their relationships, which are defined as tuples. The extension of the tuple is given by its places that can involve any kind of entities and a tuple type defines the types that the places of its instances must instantiate. For example, the wholeParts tuple represents a relation between two elements in which the 4D extent of one element is completely contained within that of another element for the entire existence of both. The example of John and his brain would fit this kind of tuple. However, the adoption of tuples where their extension is defined by the entities that take place in these tuples have been argued as rather counter-intuitive in comparison with 3D ontologies [44]. Similarly, our exploratory analysis concerning 3D and 4D ontologies [34], reported that subjects found the representation of relationships in the 4D ontology-driven models difficult to comprehend and had an unnatural feel towards defining relationships between entities. Therefore, we define our last hypothesis as follows:

- *H_3: The formation of relations between entities is more difficult to comprehend with 4D ontology-driven models than with 3D ontology-driven models, given a certain understanding of the respective ontology.*

Based upon these three hypotheses, we will discuss the experimental design of our empirical study below.

4 Experimental Design

In this section we will outline our experimental design based upon the research of Wohlin et al. [46] in order to test the hypotheses above. This experimental design was presented and discussed with a panel of experts at the 11th International Workshop on Value Modeling and Business Ontologies [47]. The feedback received was used to further fine-tune the experimental design. As such, we first define the variables to test. Next, we specify the selection of our subjects. Moreover, we explain the choice of our experimental design type, and the instruments that will be applied in this experiment. Finally, we discuss the internal validity of the experiment.

4.1 Variable development

Independent Variables

In our study, the independent or affecting variables constitute of the meta-ontological choices that ontologies make. As mentioned above, in order to emphasize these differences of meta-ontological choices, we opted for a 3D and a 4D ontology to construct the ontology-driven models, with the purpose to highlight their differences. Thus, in our experimental setting, we can control if we either assign our test subjects with the UFO (3D) or the BORO (4D) ontology-driven models.

Dependent variable

As formulated in our hypotheses and research question, we are interested in measuring the pragmatic quality or the model comprehension of these 3D and 4D ontology-driven models. The pragmatic quality of a conceptual model can be determined through the understandability or comprehension of a model by its users. As such, we will focus on the model comprehension of our conceptualizations to assess their pragmatic quality. Model comprehension can be measured with different approaches. A distinction is made between efficiency and effectiveness [48]. While effectiveness of a modeling technique is defined by how well it achieves its objective – in our case model comprehension – efficiency is defined by the effort required to apply the modeling technique. The former can be measured by output measures evaluating the quantity and/or quality of the results; the latter can be measured by a variety of input measures such as time, cost or effort. In our paper, the effectiveness will thus directly measure model comprehension, while the efficiency will measure the cost of effort to comprehend the models.

More specifically, we measure the **effectiveness** of the ontology-driven models with comprehension and problem-solving questions. These output measures are similar to previous research studies [21, 49, 50], where they also compared the comprehension and understandability of different kinds of models that were constructed with different development techniques. While the *comprehension questions* assess a basic level of model comprehension, the *problem-solving questions* are more challenging and target a deeper level of model comprehension from the subjects. More specifically, in our experiment the comprehension questions serve two purposes: first they aim to evaluate if a subject fully understood what real-world situation the ontology-driven model is representing. Second, they assess if the subject correctly interprets the underlying structure and meaning of the ontology. The problem-solving questions on the other hand assess if a subject did not only understand the model but can also apply the ontology for defining new concepts, new relations and by framing new modifications to the ontology-driven models. Since both types of questions hold only one correct answer, they can be objectively corrected. Each correct answer corresponds to one point. At the end, the total number of points can be compared to assess the number of correct answers.

The **efficiency** of the ontology-driven models will be measured by: (1) assessing the *amount of time* needed to understand the models, and (2) the amount of effort a subject had to spend to fulfill the tasks related to the ontology-driven models, here expressed as the *ease of interpretation* (EOI). Our EOI questions are based on the perceived ease of understanding as applied in several research efforts [25, 51, 52]. The EOI questions are divided in such a way that they measure different aspects of perceived effort during the experiment. More specifically, they assess: (i.) the effort in comprehending a specific assignment; (ii.) the effort spent to complete the comprehension questions or the problem-solving questions; and (iii.) which assignment required the most effort to solve.

Control Variable

Since we will be testing users' comprehension of 3D and 4D ontology-driven models, we need to be certain that all subjects have an equal understanding of the 3D or 4D models they are dealing with. Therefore, we need to assure that the interpretation of a certain model can be linked to the ontology that was applied to construct the model, and not to a limitation of the subject's knowledge of the ontology. As such, we apply a control variable to test every subject's knowledge and understanding of the ontology, before the start of the experiment. The results from the subjects that failed the knowledge test will not be incorporated into the results of the experiment.

4.2 Subject Selection

The subjects in our study all had prior education in the domain of conceptual modeling and were completing their Masters in Business Engineering at the University of Ghent. As stated by Falessi et al. [53], using students as participants remains a valid simplification of reality needed in laboratory contexts. It is an effective way to advance software engineering theories and technologies but, like any other aspect of study settings, should be carefully considered during the design, execution, interpretation, and reporting of an experiment. We decided to select students as our test subjects since they have no prior knowledge of ontologies and can thus be seen as a 'tabula rasa'. Consequently, we can train them with an ontology and a new paradigm without the interference of any pre-used paradigm of another ontology. This allows us to measure the full impact of the comprehension of the ontology-driven models. A total of 156 subjects participated in the study, of which 78 in each treatment. All subjects have the same age, with an average age of approximately 23 years old and with an equal gender balance between male and female. The majority of our subjects have a business/technical-oriented background, with prior modeling experience of the EER conceptual modeling technique and the BPMN process modeling notation, through both theoretical classes and practical sessions. The subjects received no prior training in ontologies. This specific selection thus leads to a controlled sample of subjects with the same level of experience in conceptual modeling and with no prior knowledge of the ontologies that were applied in the empirical study. To determine the minimum number of subjects for our empirical study, we based ourselves on the differences in the averages in the model comprehension scores from the study of [34]. Based upon the sample size formula below [54], assuming a Type I error (α) of 5% and a Power (1−β, where β is Type II error) of 0.8, we require a total number of 43 subjects per treatment group. Since we have a total of 156 subjects, of which 78 in each treatment, we can confirm that the number of participants in our experiment is ample with regards to the required statistical minimum.

4.3 Experimental Design Type

An experiment consists of a series of tests of different treatments [36]. To get the desired results to answer our research question, the series of tests must be carefully planned and designed. Based on our hypotheses, we can derive two treatments: a BORO treatment and a UFO treatment. The series of tests in each treatment comprises different models with each emphasizing a specific metaphysical characteristic. Our subjects were thus divided into two different treatments, where each treatment submits the subjects to similar tests where the comprehension of the models is measured. We assigned the subjects randomly to these treatments according to the *balancing design principle.* By randomizing we mean that subjects were allocated randomly to either one of the treatments. By balancing the treatments, we assigned an equal number of subjects to each separate treatment, to arrive at a balanced design. Balancing is desirable since it both simplifies and strengthens the statistical analysis of the data [36]. The design type of our experiment is a *one factor with two treatments* design, meaning that we compared the two treatments against each other with one independent variable – model comprehension. Each subject also takes part in only one treatment. Most commonly, the means of the dependent variable for each treatment are compared. We thus assigned scores to the different measures of the dependent variable – the comprehension questions, the problem-solving questions, the amount of time required to solve the task and the ease of interpretation questions – in order to compare our two different treatments objectively. This aspect will be discussed in more detail in section 4.4. As a summary of this section, Figure 2 gives a more comprehensive overview of the different aspects of our experimental design.

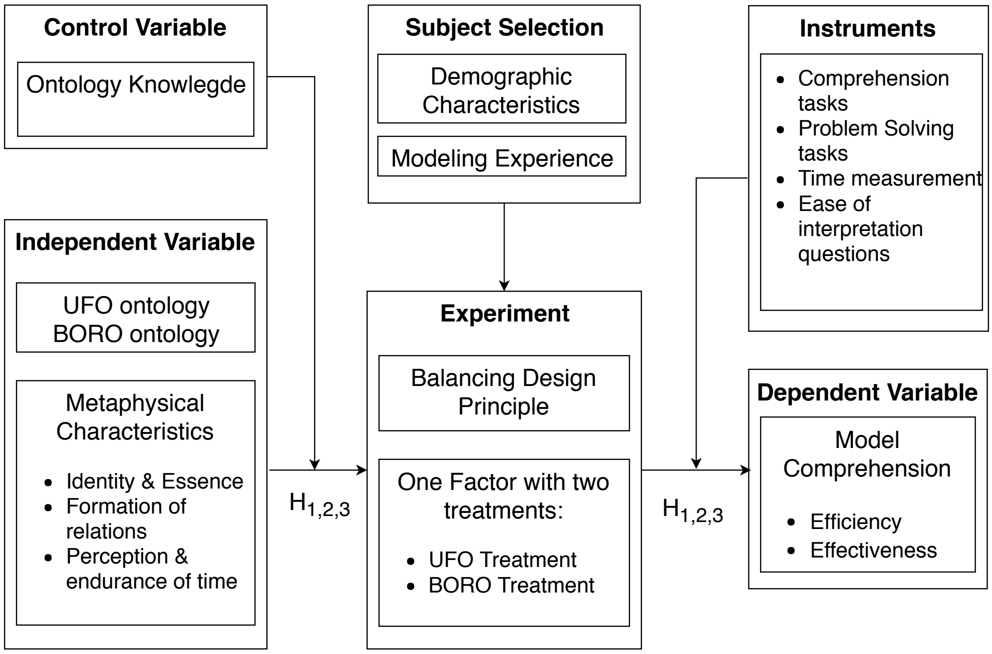

Figure 2: Overview of the Experimental Design

4.4 Instrumentation

Below, we describe in detail the different phases a subject goes through when participating in our experiment, and the kinds of instruments we applied in each of these phases. All materials – the ontology-driven models, knowledge assessments, comprehension questions etc. – of this empirical study can be found at our online repository at Open Science Framework (OSF)[4].

1. Ontology training and its modeling approach

Each subject is trained in either UFO or BORO, depending on the group or treatment they belong to. The ontology and its modeling technique are explained with the aid of a description of the ontology. Both ontologies can be expressed with the UML notation. Each modeling technique consists of a UML profile that reflects the ontological distinctions prescribed by the respective ontology. For UFO, this technique is called OntoUML [29], while BORO has the BUML modeling technique [30]. By expressing both ontologies in the UML notation, we can eliminate any errors that could occur from applying different modeling notations. Additionally,

[4] https://osf.io/ahfjc/

all our subjects received previous courses in UML modeling, making the notation quite familiar, and allowing us to fully measure the comprehension of the models resulting from applying the specific modeling technique – OntoUML or BUML – without 'interference' of the modeling notation. In the description of the ontologies, there is also a section that briefly describes the different syntactic elements of UML. Each subject could take as much time as they needed to read and understand the description. The description was drafted together with several small modeling examples in order to strengthen the subjects' understanding of the respective ontology (see OSF repository).

2. Control variable: Assessment of subjects' knowledge

To assess if the subjects clearly understood the respective ontology and corresponding modeling approach, we evaluate each subject's understanding with several written statements concerning the ontology (see OSF repository). Each of these statements describes a certain phenomenon or scenario, to which the subject has to choose the correct corresponding element of the ontology. The subjects can choose from four different multiple-choice answers, with only one correct answer. In total, ten statements were given for each treatment. All statements were derived from examples drawn from existing literature related to the ontologies. We would like to emphasize that the statements for both treatments were identical. Every subject was thus submitted to the same assessment, of course with varying answers depending on the ontology. A student failed the assessment if they were not able to answer 50% of the statements correctly. These subjects could still participate in the experiment though, so that we could assess how much impact the knowledge of an ontology had on interpreting and comprehending the ontology-driven models.

3. Interpretation of the models

After the assessment of the respective ontology, the subjects were submitted to the treatment, where they received the three assignments and their related models and questions in a sequential manner. More specifically, we developed three ontology-driven models of both the BORO [27] and UFO [29] ontologies, which our subjects were tasked to interpret (see OSF repository). Since we distinguish 3D and 4D ontologies according to their differences in metaphysical characteristics, we created each model in such a way that it emphasizes one of the metaphysical characteristics, as described in section 2. In other words, the difference between these models or conceptualizations represents how the respective ontologies deal with their metaphysical characteristics. Table 1 summarizes the metaphysical characteristic each model was focusing on and explains the scenario that was being modeled. Each of these models was presented to, and approved by, an expert of the respective ontology (*Supra:* acknowledgements). We would further like to note that each of the assignments also represent an entirely different domain in order to make the interpretation of the models more neutral towards any domain or purpose that would suit a certain ontology better. Furthermore, by using different domains for each assignment, subjects do not benefit from an increased understanding of a specific domain throughout the experiment's different assignments.

Table 1: Metaphysical Characteristics of the different models

Metaphysical Characteristic	**Model Scenario of Assignments**
The notion of identity	The first model represents a type of aircraft comprised of different kinds of components, specifically one of the components is a type of fuel pump that is part of the aircraft.
The perception and endurance of time	The second model represents a scenario where a company keeps track of its projects and where events such as the start and end of a project are recorded.
The formation of relationships between entities	The third model represents a layered composition of interconnected protocols which together form a protocol stack.

The experiment was conducted without any further explanation or description. We would like to emphasize that for every assignment, the models in UFO and BORO are informationally equivalent. They thus represent an identical scenario. Next, in order to measure model comprehension, the subjects were given a set of comprehension and problem-solving questions that were related to each model.

Comprehension Questions

After the subjects completed the interpretation of a model, they answered a set of comprehension questions to assess their interpretation. During the questions, subjects could always consult the model of the respective task. The comprehension questions were in the form of multiple-choice questions, with only one correct answer. These questions reviewed a subject's interpretation of the model and their comprehension of the concepts and structure of the ontology. All comprehension questions were the same for both the UFO assignments as for the BORO assignments (see OSP project). A score was then calculated – independently by an automatic marking system – depending on the number of correct answers given by the subject.

Problem-Solving Questions

After the subjects completed the comprehension questions, they had to answer several problem-solving questions. We would like to note that the subjects received the correct model related to the respective assignment after answering the problem-solving questions. As such, a subject would not continue the new task with any wrong assumptions made during the next problem-solving task. The problem-solving questions were also in the form of multiple-choice, with only one correct answer. Again, a score was calculated depending on the number of correct answers given by the subject.

4. Ease of Interpretation

As the last phase of the experiment, several EOI questions were asked at the end of every assignment to assess which kind of questions were perceived as most difficult (see OSF repository). The EOI questions from 1-4 were repeatedly asked after completing every assignment. As mentioned above, our EOI questions measured various aspects of perceived effort during the experiment. While EOI questions 1-3 assessed the overall perceived difficulty of the assignment, EOI question 4 measured the difference in perceived effort between solving the comprehension questions and the problem-solving questions per assignment. After the three assignments were completed, two final EOI questions (5 and 6) were asked to assess which assignment was perceived as the most cumbersome to solve and how confident they felt in solving the tasks related to the interpretation of the ontology-driven models.

4.5 Internal Validity

To avoid any threats to the internal validity, we carefully designed and monitored the experiment. Several experimental standards were also implemented to strengthen the validity of the experiment: (1) subjects were selected on a random basis, (2) we applied the *balancing design principle* in order to balance our treatments; (3) subjects were selected from a 'controlled' environment, meaning that they all shared the same background and shared similar experiences with conceptual models; (4) neither of the subjects had any prior knowledge of either of the ontologies that were applied in the treatments; (5) we inserted a control variable in the experiment to assess that subjects had a similar understanding of the ontologies before commencing the experiment; and finally (6) our experimental design was evaluated by several subjects before the actual experiment took place, in order to test for any ambiguities or dubiety in the assignments, models or questions.

5 Results of experimental study

Below we will first discuss the descriptive results related to the knowledge assessment, the effectiveness and the efficiency of each treatment. Next, we test the hypotheses as formulated above and examine if significant differences can be deducted.

5.1 Descriptive statistics

Knowledge Assessment

When we examine the results of the knowledge assessment test – which was our control variable – we can conclude that all subjects gained a reasonable understanding of the ontology's structure and concepts, with an average score of 85.58% for the BORO treatment and 91.83%

for the UFO treatment. None of the participating subjects gained a score lower than 50%, which would have excluded the results of the particular subject in the experiment.

Effectiveness of the treatments

Table 2 displays the average result scores of the individual assignments, the total score of the assignments for each treatment and the scores for both comprehension and problem-solving questions. As for the total average scores of each assignment, the table demonstrates that for assignment 1, which deals with the notion of identity, BORO scored slightly better than UFO. For assignment 2, which handles the perception of time, and assignment 3 which focuses on the formation of relations, UFO scored higher compared than BORO. Especially in assignment 2 we notice the most substantial difference between the two ontologies. Overall, when we calculate the total scores for both treatments, UFO subjects scored an average of 74.33%, compared to a 65.92% of BORO subjects. A closer look at the scores of the comprehension questions reveals that BORO scores higher in both assignments 1 and 3, where especially assignment 1 differs substantially from UFO. On the other hand, UFO scores were considerably higher in assignment 2. Further it would seem that both treatments scored the least on assignment 1, which focuses on the principle of identity. Overall, UFO (69.84%) scores slightly higher than BORO (63.83%) with regards to the total average score of the comprehension questions. Additionally, we would like to note the evolution in score results, where UFO subjects take a ‘leap’ after assignment 1, where they scored on average 30% higher compared to assignment 2 and on average 25% compared with assignment 3. In BORO, subjects tended to have the same, rather low scores in assignments 1 and 2, only improving their comprehension scores in assignment 3. When we consider the problem-solving questions, we notice that overall, UFO scored higher on every assignment compared to BORO. It would thus seem that UFO subjects had less difficulty answering the problem-solving questions compared to the BORO subjects. Regarding the difference in the total average score of the assignments, UFO (84.94%) scored substantially higher than BORO (70.51%). The difference in total score between UFO and BORO for the problem-solving questions (14%) was also more profound compared to the difference in score of the comprehension questions (6%). These results seem to suggest that BORO subjects put in reasonably more effort in answering the problem-solving questions compared to UFO subjects.

Table 2: Average scores of the experiment

Average Scores		Assignment 1	Assignment 2	Assignment 3	Total Score Assignment
Total Average scores	BORO	60.36%	59.34%	77.66%	65.92%
	UFO	58.55%	81.56%	81.14%	74.33%
Comprehension questions	BORO	54.70%	56.28%	80.51%	63.83%
	UFO	46.15%	83.76%	79.62%	69.84%
Problem-solving questions	BORO	73.08%	66.99%	70.51%	70.51%
	UFO	83.33%	76.60%	84.94%	84.94%

Efficiency of the treatments

As a first measure of the effort required to comprehend and understand the models of the respective treatments, we consider the average time needed to solve the assignments. As displayed in Table 3, subjects of the BORO treatment (40:02) required only slightly more time in solving all the assignments compared to the UFO treatment (39:07). However, this rather small difference does not entail that BORO subjects perceived the assignments as more difficult, and therefore requiring more time than the UFO treatment. This is again confirmed when looking at the average time needed to complete the individual assignments, where there exists almost no difference between the treatments in the time required to solve these assignments.

Table 3: Average amount of time needed to finish the experiment (mm:ss)

	Assignment 1	Assignment 2	Assignment 3	Total Time
BORO	16:55	11:54	11:13	40:02
UFO	16:12	11:43	11:12	39:07

Figure 3, 4 and 5 provide an overview of the number of answers given to the ease of interpretation questions. These figures display the specific answers (e.g. A to E) of all subjects on the EOI questions. As for the first EOI question, the highest number of subjects that experienced difficulty in interpreting the model, both for the UFO (37) and BORO (45) treatment, is related to the first assignment. Also, the third assignment reports a majority of subjects that perceived the task as rather difficult. Concerning the second EOI question, slightly more subjects of the UFO treatment perceived the comprehension questions of the third assignment as most difficult, compared to the BORO group where most subjects reported the comprehension questions of the first assignment as most difficult. While the perception of difficulty concerning the comprehension questions is similar between both ontologies, the difference between both ontologies appears to be the most prominent relating to the first assignment involving the notion of identity, where around 25% of the subjects indicated that they perceived the comprehension questions as rather difficult or very difficult, compared to approximately 40% of the subjects in the BORO treatment. Regarding the third EOI question, subjects indicated that the problem-solving questions were perceived as difficult. In the third assignment even 20% of the total number of subjects, again for both treatments, perceived the problem-solving questions as very difficult. It would thus seem that the problem-solving questions were perceived as more difficult compared to the comprehension questions.

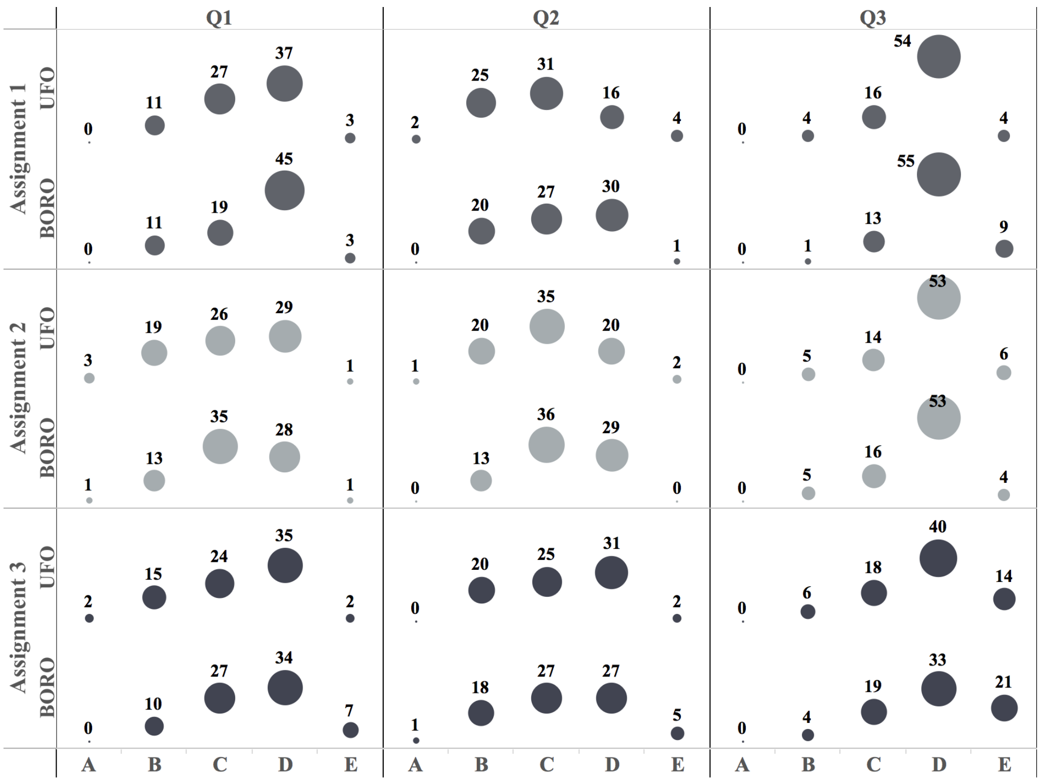

Figure 3: Results Ease of Interpretation Questions 1-3

This perception is again confirmed when we consider Figure 4 where the results of the fourth EOI question are displayed. This question clearly demonstrates that for all three assignments and for both treatments, more than 60% of the subjects perceived the problem-solving questions as most difficult to solve when compared to the comprehension questions. Interestingly the average scores of the problem-solving questions are considerably higher compared to the scores of the comprehension questions. A reason for this could be that because of their more difficult nature, a subject had to think more thoroughly on the answer of the problem-solving question, and the associated structure of the ontology, leading to a higher number of correct answers.

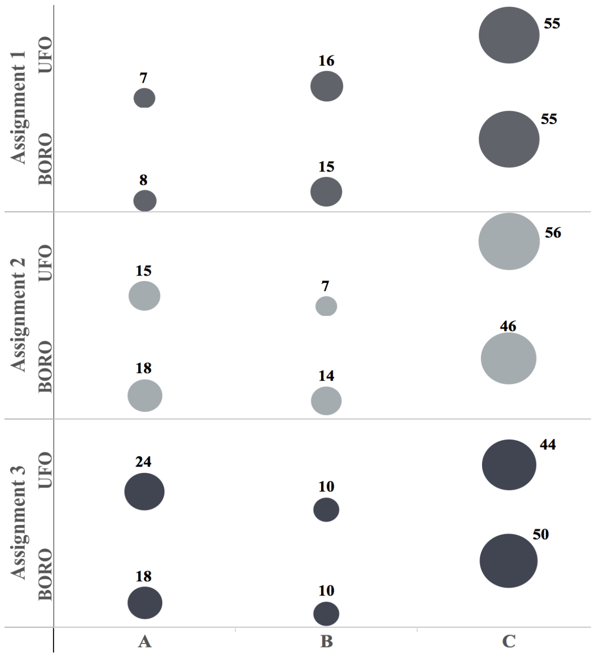

Figure 4: Results Ease of Interpretation Question 4

As for the EOI questions at the end of the experiment – represented in Figure 5 – the first question indicates that 49% of the subjects of the BORO group perceived the first assignment as the most difficult to solve, while 45% of the subjects of the UFO group experienced the third assignment as most difficult. These results mark a rather clear difference in perception between the two treatments. As for the final questions, most of the subjects in both treatments answered that they felt positive in terms of interpreting more models related to the ontology, and that they believed to reasonably understand the ontology's structure and concepts. We can also note that reasonably more subjects of the BORO treatment (31%) compared to the UFO treatment (15%) answered this question negatively, signaling that the concepts and structure of the ontology were still vague to them. This perception also corresponds to the total average scores as discussed above, where the UFO scores are higher than the BORO scores.

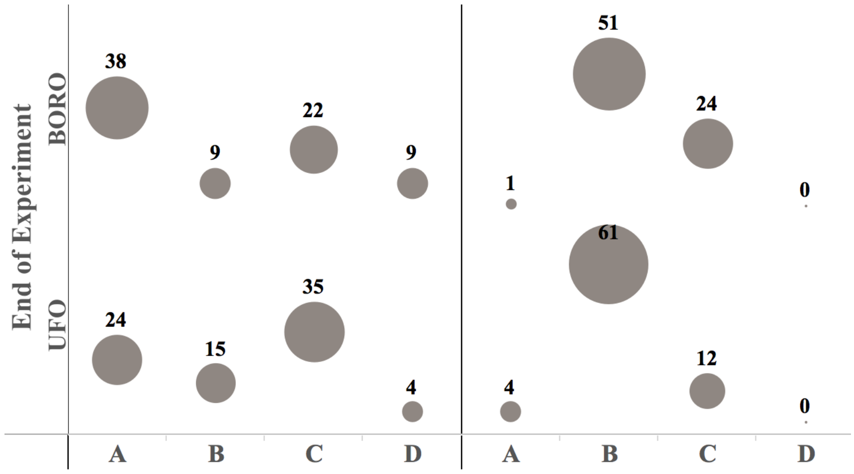

Figure 5: Results Ease of Interpretation Questions 5-6

5.2 Hypotheses Testing

Effectiveness of the treatments

To test our hypotheses, we test if the scores of each assignment between the two treatments differ significantly. To determine which kind of test to apply, we first examined the distributions of our data – the total individual scores per subject, for each individual assignment. To determine if our data is normally distributed, we performed the Shapiro-Wilk test, revealing that both the data of the BORO and UFO treatments follow a non-normal distribution (p-value: 0.000) – indicating that the hypotheses must be analyzed with non-parametric tests.

Since our data is not normally distributed and consists out of two independent groups, with no relationship between the observations in each group or between the groups themselves, our data satisfies the assumptions of the Mann-Whitney U test [55]. In Table 4 and Table 5 we have displayed the results related to the Mann-Whitney U test. While Table 4 expresses the mean ranks and the sum of ranks for each assignment for both the UFO and BORO treatments, Table 5 displays the outcome of the test and the associated p-values. The hypotheses are tested at the 95% confidence interval. Additionally, since the hypotheses are directional we adopt the one-tailed significance level.

As for the first assignment, which is related to the first hypothesis, there is a higher mean rank for the BORO treatment (82.86) compared to the UFO treatment (74.14). This difference in mean rank seems to be in line with the claim of the first hypothesis. When conducting the Mann-Whitney U test, we retrieved a p-value of 0.133 – meaning that no significant difference can be acknowledged on the 95% confidence interval between the scores of the UFO and BORO groups. In other words, we **reject H_1**, and therefore cannot confirm that the notion of identity is more difficult to comprehend with 3D ontology-driven models than with 4D ontology-driven models on the 5% significance level.

The second assignment produces the mean rank of 49.91 for the BORO treatment, compared to a mean rank of 108.09 for the UFO treatment. These mean ranks are the opposite to the claim of our second hypothesis. The Mann-Whitney U test produces a p-value of 0.000; meaning that there is a significant difference between the UFO scores and the BORO scores, however in the direction that the UFO scores are significantly higher than the BORO scores at a 95% confidence interval. We thus **reject H_2** and cannot attain that the perception of time is more difficult to comprehend with 3D ontology-driven models than with 4D ontology-driven models. In fact, the opposite is ascertained.

Finally, concerning the third assignment, we notice a mean rank of 71.7 for the BORO treatment, compared to a mean rank of 85.3 for the UFO treatment. These ranks are in line with the assumption of the third hypothesis. The Mann-Whitney U test produces a p-value of 0.028, confirming the third hypothesis at the 95% confidence interval. In other words, we **accept H_3**, and confirm that the formation of relations between entities is more difficult to comprehend with 4D ontology-driven models than with 3D ontology-driven models

Table 4: Mann-Whitney U Ranks of Scores Assignments

Ranks	Treatment	Mean Rank	Sum of Ranks
Assignment 1	BORO	82.86	6463
	UFO	74.14	5783
Assignment 2	BORO	48.91	3815
	UFO	108.09	8431
Assignment 3	BORO	71.7	5592.5
	UFO	85.3	6653.5

Table 5: Mann-Whitney U Test of Scores Assignments

Test Statistics	Assignment 1	Assignment 2	Assignment 3
Mann-Whitney U	2702	734	2511.5
Z	-1.212	-8.217	-1.916
Asymp. Sig. (2-tailed)	0.226	0.000	0.055
Asymp. Sig. (1-tailed)	0.113	0.000	0.028

Efficiency of the treatments

Similar to the hypothesis testing above, we compared if the perceived effort to comprehend the models of each assignment differs significantly between the two treatments. We thus investigate if there exist significant differences in the time needed to complete the assignments, and the answers with respect to the EOI questions. Similarly, we examined the distribution of our data with the Shapiro-Wilk test, revealing again that our data – time required to complete the assignments and the EOI results – is non-normally distributed, meaning that we can apply the non-parametric Mann-Whitney U test.

In Table 6 we have displayed the results of the Mean-Whitney U test related to the time needed to complete the assignment per treatment, while Table 7 displays the Mean-Whitney U results for the EOI answers for each treatment. For the EOI results, we grouped EOI questions one to three together per assignment, since these questions aim to assess the difficulty of the respective assignment. The fourth EOI question and the two final EOI questions (5-6) asked at the end of the assignment were tested separately since they each measure different aspects of perceived effort as explained above.

In line with our observations when presenting the descriptive statistics, we notice no significant differences in the time required to complete the assignments between the two treatments. As for the results concerning the EOI questions, we notice a difference in answers from the EOI questions assessing the difficulty of assignment 1, and for the fifth EOI question that aims to assess which assignment was considered as most difficult. As we observed in the descriptive statistics from our EOI questions, subjects from the BORO treatment indicated that the first assignment – related to the notion of identity – was more difficult to comprehend, while the UFO treatment indicated that the third assignment – related to the formation of relations between entities – was more difficult to support. The differences between these perceptions thus seem to be significant, supporting the rejection of H_1 and the acceptance of H_3.

Table 6: Mann-Whitney U Test of Time per treatment

Test Statistics	Assignment 1	Assignment 2	Assignment 3
Mann-Whitney U	2814	3023	2979
Z	-0.808	-0.067	-0.223
Asymp. Sig. (2-tailed)	0.419	0.946	0.823

Table 7: Mann-Whitney U Test of EOI results

Test Statistics	EOI Assign. 1	EOI Assign. 2	EOI Assign. 3	EOI-4	EOI - 5	EOI - 6
Mann-Whitney U	2549.5	2730.5	2716.5	2996	2463.5	2666.5
Z	-1.971	-1.202	-1.247	-0.202	-2.609	-1.415
Asymp. Sig. (2-tailed)	0.049	0.229	0.212	0.84	0.009	0.157

6 Protocol Analysis

In the first phase of this empirical study, we conducted an experiment with the purpose of generating a sufficient amount of data to test our hypotheses. However, in order to provide additional insights into the nature of our results, we then conducted a more in-depth analysis – in the form of a protocol analysis. While the experiment was performed on a larger scale, the protocol analysis was performed with a smaller set of fewer subjects, since the goal of the protocol analysis is not to produce data, but to acquire knowledge on how subjects perceived the experiment [37–39].

6.1 Design of Protocol Analysis

A protocol analysis is a research method that elicits verbal reports from research participants, which reveals the mental processes taking place as individuals work on the interpretation of the models. Subjects are required to verbalize their thought processes and strategies, as well as to verbalize their answers to the comprehension, problem solving and EOI questions. These verbal reports and the progress of the subjects were closely monitored by a researcher guiding the treatment. Hence, we performed the protocol analysis on a new set of subjects, in the exact same way as our experiment, but with the sole purpose to better understand the outcome of our results. By performing this protocol analysis, we were able to observe in which phases of the experiment subjects experienced any difficulties, allowing us to better comprehend how the subjects in the experiment perceived the assignments. In line with other protocol analysis studies [37, 56], the number of subjects participating in the protocol study was small, a total of 6 participants. Similar to the experiment, the subjects all had prior experience in the domain of conceptual modeling and no prior knowledge of ontologies. The subjects were evenly distributed over the two treatments, since a large volume of data is generated even with a small sample size.

6.2 Results of Protocol Analysis

Training of the ontology

After reading the UFO description, most subjects reported difficulty in understanding the Moments aspect of the ontology, where the concepts Relators, Modes and Qualities were thought of as rather troublesome to comprehend. Furthermore, the structure of the overall ontology, and the interrelations between the ontological concepts were described as confusing. This is much in contrast to the BORO ontology, where subjects reported the outline and concepts as very comprehensible. This ease of comprehension can probably be assigned to the fact that BORO defines only a few concepts and relations.

Assignment 1

As for the first assignment, UFO subjects frequently reported difficulty in differentiating between Kinds, SubKinds, Collections and Categories. Furthermore, answering the comprehension questions required a substantial amount of the time to solve the first assignment. Especially the comprehension questions that focused on assigning the ontological concept to a class diagram were perceived as the most demanding kinds of questions. Additionally, the statements concerning the actual representation of the model were perceived as difficult, especially question 7 focusing on types and instantiations, where all subjects required more time to solve the question in comparison with similar questions in the set. Compared to the comprehension questions, the problem-solving questions were perceived as more feasible. When the subjects were able to see the actual ontological concepts that were linked to the class diagrams, it was easier for them to create and instantiate new ontological concepts. This corresponds to the descriptive results of the experiment, where the total average score of the problem-solving questions (83.33%) was considerably higher compared to the total average score of the comprehension questions (46.15%).

As for BORO, subjects clearly perceived the first assignment as more challenging than expected after going through the description. While the differentiation between Tuples, Types and IndividualTypes was reported as clear during the reading of the description, applying the concepts in the assignment was less straightforward. When answering the comprehension questions, subjects expressed a high sense of doubt on the correctness of their answers. Especially the distinction between Types and IndividualTypes in the model was hard to differentiate. This perceived difficulty of the principles of identity and essence-defining properties in BORO explains why the first hypothesis was rejected in the experiment. Next, similar to UFO, the problem-solving questions were perceived as more doable. Again, this is in line with the results of the experiment, where the total average score of the comprehension questions (60.36%) was substantially lower compared to the problem-solving questions (73.08%). Regarding the EOI questions, we see the opposite, where substantially more subjects of both the UFO and BORO treatment ranked the problem-solving questions as more difficult than the comprehension questions.

Assignment 2

When submitted to the second assignment, subjects of the UFO treatment reported both comprehension and the problem-solving questions as easier to answer compared to the previous assignment. While there still existed some doubt in assigning the correct ontological concepts to the corresponding class diagrams, it clearly required less effort than before. Almost none of the subjects reported any substantial difficulties when answering the comprehension questions related to the second assignment. Also, interpreting the ontological concepts related to time presented no real difficulty. As for the problem-solving questions, they were deemed somewhat more challenging. With these questions, subjects did mention that applying the time related concepts (e.g., quality) in new scenarios required more effort to solve. The perception of our subjects in the protocol analysis concerning this assignment also matched the descriptive results from the EOI questions of the experiment, where more subjects indicated that they perceived the second assignment as easier to solve compared to the rankings of the first assignment. Likewise, for the results of this assignment, the average total scores greatly increased compared to the first assignment, with a total average of 83.76% for the comprehension questions and 76.60% for the problem-solving questions.

As with the UFO subjects, the participants of the BORO treatment described the second assignment as considerably easier compared to the first assignment. Subjects required less time to solve the assignment and reported fewer difficulties when answering both sets of questions. Several comments however were made concerning the happensIn TupleType, which is used to represent an elements relationship with a protion of space-time, and its relation to the wholePart TupleType and the Event IndividualType. Their specific relation was not always clear. Nonetheless, after finishing the comprehension questions, subjects did describe the comprehension questions as more feasible than the previous assignment, and they believed they had assigned the correct ontological concepts to the different class diagrams in the model.

However, when given the correct model in the problem-solving questions, subjects noticed that many of their answers were not correct and that they had erroneously assigned the ontological concepts. This is an important observation since it seems that although subjects were convinced that they correctly interpreted the concepts underlying the BORO ontology, they had not yet truly understood the structure of the ontology. Most of these incorrect answers were associated with interpreting the happensIn TupleType and the wholePart TupleType, which caused some initial doubt when answering the comprehension questions. This observation accurately explains the rejection of our second hypothesis. Similarly, to the experiment, the protocol analysis also designates that the perception of time appears to be more difficult to comprehend with 4D ontology-driven models than with 3D ontology-driven models.

With regards to the problem-solving questions, these were reported to be more difficult than the comprehension questions, with the main reason being that applying these concepts in new situations related to the model was more difficult. In the experiment, subjects received a total average of 56.28% for the comprehension questions and 66.99% for the problem-solving questions. When we interpreted these results with the perceptions of the protocol analysis, it appeared that subjects had the misconception of properly understanding the concepts and structure of the BORO ontology, while the actual results indicate otherwise. The results of the EOI questions seem to confirm this assertion, since the second assignment was rated as the least difficult out of all three assignments.

Assignment 3

In the last assignment, subjects of the UFO treatment displayed little difficulty in assigning the correct ontological concepts to the respective class diagrams. The distinction between identity-based concepts such as Kinds, Roles, SubKinds, etc. was reported as clear. Also the Relators representing the relationships in UFO could be easily recognized. Several subjects mentioned that the previous assignments had significantly increased their insights in the structure of the UFO ontology. Instead, subjects struggled more when answering the statements of the comprehension questions, which assess their interpretation of the model. As for the problem-solving questions, these were reported as rather more difficult compared to the comprehension questions. Most of the effort to solve the assignment went into answering the problem-solving questions. Subjects mentioned that they understood the UFO concepts and relations, but that applying the ontology in new situations was still challenging. Most of the doubt originated from assigning the mediation and material relationships in a Relator. Despite being reported as more difficult, subjects did answer many of the problem-solving questions correctly. The perceptions in our protocol study aligned with the results of the experiment, where both the comprehension (79.62%) and the problem-solving questions (84.94%) gathered rather high scores. Equivalently, the subjects of the BORO group experienced fewer problems in assigning the ontological concepts to the class diagrams compared to the previous models. Similarly, subjects commented that the practice of the previous models helped in identifying the correct ontological concepts in the current assignment. When answering the comprehension questions related to the interpretation of the model, almost all subjects mentioned experiencing problems in identifying which was the client protocol and which the supplier protocol. For clarification, we have represented a fragment of both the BORO and UFO protocol model in Figure 6 to illustrate the difference between the two. The assignment corresponding to this figure represents a communication protocol that forms a layer in a protocol stack, where a client and a supplier protocol indicate which position, they take in the protocol stack. An instance of this representation is the Universal Mobile Telecommunications System (UMTS) protocol stack, where the media access control (MAC) and Radio Link Control (RLC) are communication protocols that form separate layers of UMTS – MAC being the supplier protocol and RLC the client protocol. In the BORO model – panel B of the figure – the distinction between the client and supplier protocol is rendered through a TupleType, where the client and supplier each take a place in this Tuple. In contrast, UFO represents these client and supplier protocols as separate classes instead of relations, probably making it easier for UFO subjects to identify and distinguish between these kinds of protocols. It is perhaps due to such differences in representation between both ontologies that the formation of relations between entities is more difficult to comprehend with 4D ontology-driven models than with 3D ontology-driven models – leading to the acceptance of our third hypothesis.

Next, the problem-solving questions of this assignment were recognized to be rather difficult. Subjects especially reported the differentiation in BORO between the place1Type or place2Type and the tuplePlace1 or tuplePlace2 relations confusing. These different relations are made to differentiate between the type level and instantiation level of Tuple relations in BORO. When we consider the experiment results, we notice a similar observation, in the sense that the scores of the problem-solving questions (70.51%) are considerably lower than those of the comprehension questions (80.51%).

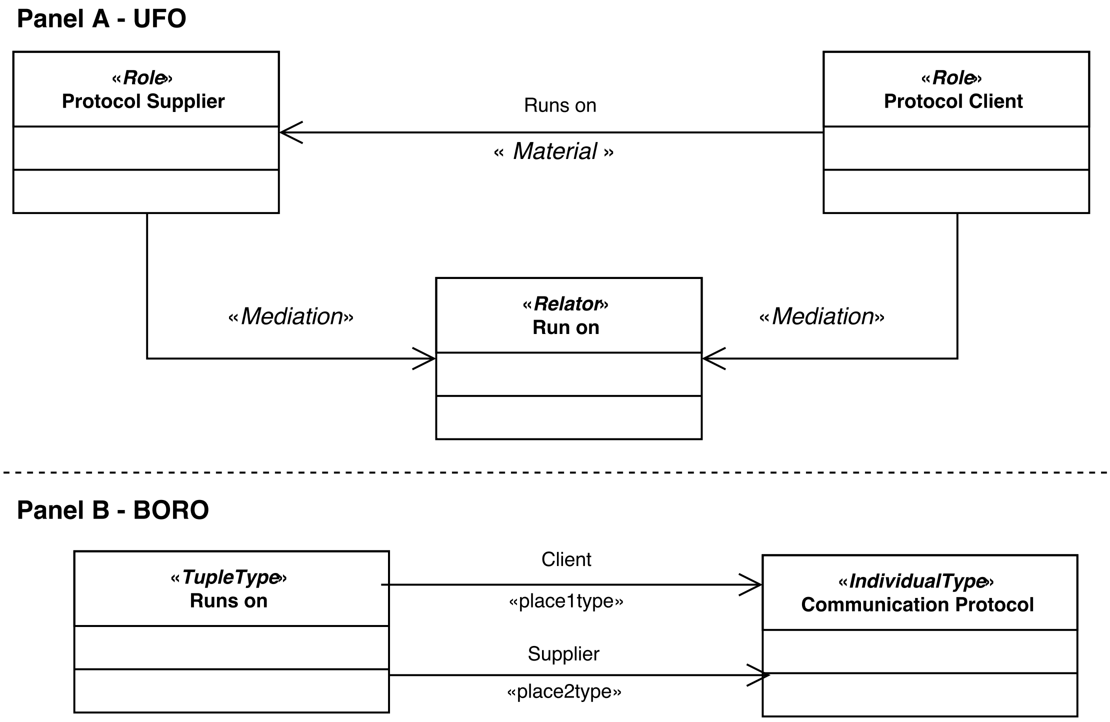

Figure 6: Fragments of UFO and BORO models in representing protocols

7 Discussion

Our earlier research efforts [33, 34] acknowledged that the conceptualizations realized by different ontologies can have a considerable impact on their pragmatic quality, but they still left us with the question – the research question of this article – to what extent is the pragmatic quality of these ontology-driven models influenced by the choice of a particular ontology. The results of our empirical study now confirm that the choice of an ontology can lead to significant differences in subjects correctly interpreting and comprehending the model (effectiveness), as well as in their perception or their effort required to comprehend these models (efficiency). In this section, we will summarize several derivations to which we can attribute these variations in model comprehension.

Derivation 1: *The paradigm underlying a 4D ontology is more difficult to comprehend than the paradigm of a 3D ontology.* Our first hypothesis assumed that the notion of identity would be more difficult to comprehend with 3D ontology-driven models than with 4D ontology-driven models. Although the total average scores of the 4D treatment were higher than those of the 3D treatment, the experimental results were not significant. Even more, the assignment related to the metaphysical characteristic of identity was perceived as the most difficult to comprehend by the 4D treatment group. During the protocol analysis, it was observed that while initially subjects of the 4D treatment did report the structure, concepts and identity principles of the 4D ontology as simple and easy to understand, this perception was rapidly disproven when the subjects were tasked to interpret the models. Even in the last assignment, several subjects remained confused about correctly identifying and distinguishing the 4D ontology concepts. Further questioning during the protocol analysis indicated that the paradigm underlying the 4D ontology seemed to be the root of this confusion. Viewing individual objects as four-dimensional, while being composed of spatial and temporal parts often left subjects disoriented. Especially whole-part relationships that are formed between such four-dimensional objects were described as counterintuitive. It would thus seem that the disadvantage of the counterintuitive paradigm, as also noted by [44], has a greater impact on comprehending the resulting models than expected. Similarly, however, also subjects of the 3D treatment encountered problems with the identity principles that are related to the ontology. In this case, it was not the paradigm that caused these problems. Instead, during the protocol analysis, subjects reported the multitude of concepts and the distinctions between the principles of identity as troublesome. Notwithstanding these initial struggles, as soon as subjects grasped the exact distinction between the different concepts, no further issues were noted in later assignments. The paradigm of the 3D ontology thus seems easier to understand and comprehend than the extensionalist paradigm associated with the 4D ontology.

Derivation 2: *The notion of time is easier to comprehend with 3D ontology-driven models than with 4D ontology-driven models.* As for the second hypothesis, related to the metaphysical characteristic of how an ontology dealt with time, it was expected that its perception would be easier to comprehend with 4D ontology-driven models than with 3D ontology-driven models. However, our results proved to be quite the contrary. While the subjects of the 3D treatment achieved their highest scores on the time-related assignment, subjects of the 4D treatment ranked lowest. Our results also proved to be significant, in the way that the perception of time is easier to interpret with 3D ontology-driven models compared to 4D ontology-driven models. Again, the immutability of objects in space-time seemed to complicate matters for the subjects in the 4D treatment. The more 'presently-focused' paradigm of the 3D ontology, where objects are viewed only from the present and with the assumption that the same object can exist over time, appears to be more comprehensible and intuitive. It would appear that the diachronic identity aspect of 3D ontologies [57], reducing the identification of essential properties that hold over some period of time, has a less prominent disadvantage in interpreting 3D ontology-driven models. Probably, the overall 3D paradigm corresponds more to our everyday way of thinking, whereas the 4D paradigm, with its emphasis on space and time extensions, immutability and whole-part relations, feels less natural.

Derivation 3: *The formation of relationships is easier to comprehend with 3D ontology-driven models than with 4D ontology-driven models.* Our last hypothesis focused on the metaphysical characteristic of forming relations, which were presumed to be easier to comprehend with 3D ontology-driven models than with 4D ontology-driven models. Similar to the results of [34], the hypothesis was also confirmed by our results where the subjects of the 3D treatment scored significantly higher than the subjects of the 4D ontology. During the protocol analysis, subjects of the 4D treatment commented that the structure of the relationships in the model made it difficult to relate to certain concepts. Subjects of the 3D ontology often judged the relationship aspect of the ontology as the most difficult to fully comprehend, especially at the beginning of the experiment. Despite the reported difficulty, subjects could correctly associate the meaning of the relationships and the interacting elements. However, while overall scores on the assignment were considerably high, subjects of the 3D ontology perceived this assignment as the most difficult to solve.

Derivation 4: *The effort to comprehend an ontology-driven model varies substantially between 3D and 4D ontologies, and additionally, can vary heavily depending on the metaphysical characteristics of the ontology. Furthermore, a misalignment between the perceptions of a model often does not match the actual interpretation of this model.*

Although the assigned models described the same scenario for each treatment, the effort required to comprehend these models varied greatly between the ontologies. We observed a significant difference between subjects of the 3D treatment ranking the first assignment as the most difficult, while subjects of the 4D treatment ranked the third assignment as most troublesome to solve. Furthermore, depending on the metaphysical characteristic that the assignment was related to, we noticed differences in perception of difficulty, also between comprehension and problem-solving questions. The most surprising element however, is the misalignment between a subject's perception of the model and the actual correctness of its comprehension. As mentioned in the derivation above, while overall scores were considerably high, subjects of the 3D treatment perceived the assignment as the most difficult to solve. This mismatch between the perception of difficulty or ease of solving an assignment and the actual results that were achieved is a recurrent observation in our experiment. While there were more subjects of the 3D treatment that ranked the first assignment as easy compared to the subjects of the 4D ontology, the scores were lower for the 3D treatment than for the 4D treatment. Especially the comprehension questions were found to be easier by the 3D treatment, while they received substantially lower scores on this aspect of the assignment. Perhaps the greatest disproportion between perception and results was the assignment related to the time perspective, where 4D subjects perceived the assignment as the easiest to solve, while their scores were the lowest of all assignments. While 3D subjects seem to perceive some assignments or questions as difficult to very difficult, often their test scores indicate the opposite, with high-test scores for these assignments or questions. Conversely, the 4D subjects perceived questions or assignments as easy, while their test scores were rather low. This kind of misconception indicates that 4D subjects probably still lacked a true understanding of the ontology's concepts and their associated meaning.

Derivation 5: *A deep level understanding is more rapidly attained with 3D ontologies than with 4D ontologies.* As our last observation, we noticed that the test results for the problem-solving questions are consistently higher for the 3D treatment than for the 4D treatment. Since the problem-solving questions aim to assess a deeper level of understanding, these results suggest that the subjects of the 3D treatment possessed a deeper understanding of the ontology compared to the subjects' knowledge of the 4D ontology. This is contrary to the initial reports when subjects completed the description of the specific ontology. While subjects of the 3D treatment often commented that the ontology's structure is extensive and somewhat complicated, subjects of the 4D treatment on the contrary described the ontology as easy and accessible to comprehend. It would thus seem that the 3D ontology initially intimidates a user with its plethora of concepts and relations, but that a full understanding of the ontology is rather rapidly achieved. On the opposite, the 4D ontology gives a favorable first impression with its simple structure and few concepts but does not easily facilitate a deeper level of understanding. To validate this observation, we performed an additional test, where we have compared the total results for all three assignments of the problem-solving questions between the BORO and the UFO treatment. The results can be found in Table 8 and Table 9. The results confirm our observation, where the UFO treatment has significant higher scores for the problem-solving questions compared to the BORO treatment, indicating that UFO subjects attained a deeper understanding compared to the BORO subjects.

Table 8: Mean-Whitney U Ranks of Total Score Problem-Solving questions

Ranks	Treatment	Mean Rank	Sum of Ranks
Total	BORO	60.43	4713.5
	UFO	96.57	7532.5

Table 9: Mann-Whitney U Test of Total Score Problem-Solving questions

Test Statistics	Total Score Problem-Solving
Mann-Whitney U	1632.5
Z	-5.069
Asymp. Sig. (2-tailed)	0.000
Asymp. Sig. (1-tailed)	0.000

8 Conclusion

This paper presented an empirical study that investigated the influence of an ontology on the interpretation and understanding of the resulting conceptual models. More specifically, we asked ourselves the question to which degree the pragmatic quality of ontology-driven models is influenced by the choice of ontology, given a certain understanding of this ontology.

This paper contributes to the domain of ODCM by demonstrating that there exist significant differences in the interpretation of ontology-driven conceptual models that were developed by applying different foundational ontologies. Moreover, the empirical nature of this research article aims to contribute to the lack of empirical research in ODCM [16, 17]. We have linked the pragmatic quality of these conceptual models to the metaphysical characteristics of an ontology. In other words, the metaphysical characteristics determine the quality of the conceptualizations. Consequently, researchers or practitioners should be aware of the impact of adopting a specific ontology when developing conceptual models and the influence they can have on the interpretation of its users. Even more, while a 3D approach is more commonly taught for instance at universities, the different effects on the outcome and comprehension of conceptual models due to applying other paradigms (such as the perdurantist paradigm) should be taken into account when educating students. We aspire that by providing greater insights into the importance of this choice, we will enable researchers to better motivate why certain ontologies are adopted. Since this research effort is the first that empirically investigates the differences between applying two ontologies, we hope that this effort will encourage new studies to further investigate and compare new types or kinds of ontologies.

External Validity

Concerning external validity, the authors would like to acknowledge that by conducting our experiment on students, we limit the overall generalizability of our results. However, as stated by [53], using students as participants remains a valid simplification of reality needed in laboratory contexts. It is an effective way to advance software engineering theories and technologies but, like any other aspect of study settings, it should be carefully considered during the design, execution, interpretation, and reporting of an experiment. Consequently, we decided to select students as our test subjects since it was the purpose of this study to compare a 3D and a 4D ontology in a 'tabula rasa' environment – meaning that we did not want any of our subjects to encompass previous knowledge concerning the ontology. Hence, the profile of students is well fitted to the nature of such an experiment. Furthermore, we carefully aimed to avoid our tasks or assignments in the experiments to be unrealistically simple. Each of our assignments was rendered to represent real-world scenarios and systems (e.g. protocol representation). Next, in order to reduce the learning effect as much as possible, we choose a completely different domain for each of the three assignments given to our subjects. Additionally, subjects were not shown their mistakes on the comprehension or the problem-solving questions. Finally, we would like to note that we have compared ontologies by means of their metaphysical characteristics. Another point of view would be to relate them to their specific purpose or usage. In other words, although the models developed with the BORO ontology were found to be more difficult to interpret in certain cases compared to the UFO models, we should take into account that perhaps the BORO models were not meant to be easily interpreted in order to facilitate, for instance, complex re-engineering purposes. We aimed to minimize this effect by relating each of the assignments to an entirely different domain (e.g. fuel pumps corresponding to an aircraft, a project scheduler and a communications protocol stack) as to compensate for any benefits one of the ontologies could have towards a certain domain or purpose. In this respect, we would also like to note that due to the design of this experiment, and more specifically the inexperience of the subjects concerning ontologies (supra Section 4.2), we have only discussed the metaphysical characteristics of 3D and 4D ontologies to a certain degree. For instance, the notion of identity was only examined in terms of how both ontologies deal with the states, phases and roles that a certain entity can adopt or instantiate. We have not dealt with the distinctions on how identity principles deal with change, transworld identity (or possible worlds) and which principles (or properties) are to be considered essential or non-essential for an entity. We believe that a more thorough comparison and discussion between the metaphysical characteristics of 3D and 4D ontologies also provides a captivating opportunity for future research efforts – both theoretical as well as empirical.

Acknowledgements

We would like to express our sincere gratitude to Chris Partridge, of the BORO Solutions Group, and Maria das Graças da Silva Teixeira, of the Ontology and Conceptual Modeling Research Group (NEMO), for their detailed reading and revisions of the respective BORO and UFO ontology-driven models applied in this study.

Bibliography

1. Mylopoulos, J.: Conceptual modeling and telos. In: Loucopoulos, P. and Zicari, R. (eds.) Conceptual Modelling, Databases and CASE: An Integrated View of Information Systems Development. Wiley (1992).
2. Dwivedi, Y.K., Wastell, D., Laumer, S., Henriksen, H.Z., Myers, M.D., Bunker, D., Elbanna, A., M.N., R., Srivastava, S.C.: Research on information systems failures and successes: status update and future directions. Information Systems Frontiers. 17 (1), 143–157 (2015).

3. Santos, P., Almeida, J.P. a., Guizzardi, G.: An ontology-based analysis and semantics for organizational structure modeling in the ARIS method. Information Systems. 38, 690–708 (2013).
4. Beydoun, G., Low, G., García-Sánchez, F., Valencia-García, R., Martínez-Béjar, R.: Identification of ontologies to support information systems development. Information Systems. 46, 45–60 (2014). https://doi.org/10.1016/j.is.2014.05.002.
5. Endres, A., Rombach, H.D.: A handbook of software and systems engineering: Empirical observations, laws, and theories. Pearson Education (2003).
6. Muskens, J., Bril, R.J., Chaudron, M.R. V: Generalizing consistency checking between software views. Proceedings - 5th Working IEEE/IFIP Conference on Software Architecture, WICSA 2005. 2005, 169–180 (2005). https://doi.org/10.1109/WICSA.2005.37.
7. Guizzardi, Pires, L.F., Sinderen, M. Van: On the role of domain ontologies in the design of domain-specific visual modeling langages. In: OOPSLA Workshop on Domain-Specific Modeling Languages. pp. 25–38 (2002).
8. Guarino, N.: Formal ontology and information systems. In: FOIS conference. pp. 3–15 (1998).
9. Guizzardi, G., Wagner, G., Guarino, N., van Sinderen, M.: An Ontologically Well-Founded Profile for UML Conceptual Models. Advanced Information Systems Engineering. 112–126 (2004). https://doi.org/10.1007/b98058.
10. Hales, S.D.S., Johnson, T.T.A.: Endurantism, perdurantism and special relativity. The Philosophical Quarterly. 53, 524–539 (2003). https://doi.org/10.1111/1467-9213.00329.
11. Verdonck, Gailly, F.: Insights on the Use and Application of Ontology and Conceptual Modeling Languages in Ontology-Driven Conceptual Modeling. In: Conceptual Modeling - ER 2016, Lecture Notes in Computer Science, vol. 9974. pp. 83–97 (2016). https://doi.org/DOI: 10.1007/978-3-319-46397-1_7.
12. Al Debei, M.M., Al-Debei, M.M., Asswad, M.M. Al, de Cesare, S., Lycett, M.: Conceptual Modelling and the Quality of Ontologies : Endurantism Vs Perdurantism. International Journal of Database Management Systems. 4, 1–19 (2012). https://doi.org/10.5121/ijdms.2012.4301.
13. De Cesare, S., Henderson-Sellers, B., Partridge, C., Lycett, M.: Improving Model Quality Through Foundational Ontologies: Two Contrasting Approaches to the Representation of Roles. In: ER 2015. pp. 304–314 (2015). https://doi.org/10.1007/978-3-319-25747-1_30.
14. De Cesare, S., Geerts, G.L.: Toward a perdurantist ontology of contracts. Lecture Notes in Business Information Processing. 112 LNBIP, 85–96 (2012). https://doi.org/10.1007/978-3-642-31069-0_7.
15. Hadar, I., Soffer, P.: Variations in Conceptual Modeling. Journal of the Association for Information Systems. 7, 568–592 (2006).
16. Moody, D.L.: Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. Data & Knowledge Engineering. 55, 243–276 (2005). https://doi.org/10.1016/j.datak.2004.12.005.
17. Verdonck, Gailly, F., De Cesare, S., Poels, G.: Ontology-driven conceptual modeling: A systematic literature mapping and review. Applied Ontology. 10, 197–227 (2015). https://doi.org/10.3233/AO-150154.
18. Lindland, O.I., Sindre, G., Solvberg, A.: Understanding Quality in Conceptual Modeling. Ieee Software. 11, 42–49 (1994).
19. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.S.: Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. Applied Ontology. 10, 259–271 (2015). https://doi.org/10.3233/AO-150157.
20. Uschold, M., Gruninger, M.: Ontologies: Principles, methods and applications. The knowledge engineering review. 11, 93–116 (1996).
21. Gemino, A., Wand, Y.: Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. Data & Knowledge Engineering. 55, 301–326 (2005). https://doi.org/10.1016/j.datak.2004.12.009.
22. Opdahl, A.L., Berio, G., Harzallah, M., Matulevičius, R.: An ontology for enterprise and information systems modelling. Applied Ontology. 7, 49–92 (2012). https://doi.org/10.3233/AO-2011-0101.
23. Evermann, J., Wand, Y.: Toward formalizing domain modeling semantics in language syntax. IEEE Transactions on Software Engineering. 31, 21–37 (2005).
24. Guizzardi, G.: Ontological patterns, anti-patterns and pattern languages for next-generation conceptual modeling. In: International Conference on Conceptual Modeling. pp. 13–27 (2014).

25. Verdonck, M., Gailly, F., Pergl, R., Guizzardi, G., Martins, B., Pastor, O.: Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study. Information Systems. 81, 92–103 (2019). https://doi.org/10.1016/j.is.2018.11.009.
26. Guizzardi, G., Wagner, G.: Can BPMN Be Used for Making Simulation Models ? In: Barjis, J., Eldabi, T., and Gupta, A. (eds.) International Conference on Advanced Information Systems Engineering (CAiSE 2011) LNBIP vol. 88. pp. 100–115 (2011). https://doi.org/10.1007/978-3-642-24175-8_8.
27. Partridge, C.: Business Objects: Re-engineering for Reuse. Butterworth-Heinemann (2005).
28. de Cesare, S., Partridge, C.: BORO as a Foundation to Enterprise Ontology. Journal of Information Systems. 30, 83–112 (2016). https://doi.org/10.2308/isys-51428.
29. Guizzardi, G.: Ontological Foundations for Structural Conceptual Models. CTIT, Centre for Telematics and Information Technology (2005). https://doi.org/10.1007/978-3-642-31095-9_45.
30. de Cesare, S., Partridge, C.: BORO as a Foundation to Enterprise Ontology. Journal of Information Systems. 30, 83–112 (2016). https://doi.org/10.2308/isys-51428.
31. Guizzardi, G., Wagner, G., de Almeida Falbo, R., Guizzardi, RenataS.S., Almeida, JoãoPauloA., Falbo, R.D.A.: Towards Ontological Foundations for the Conceptual Modeling of Events Background : The Unified Foundational Ontology (UFO). Conceptual Modeling. 8217, 327–341 (2013). https://doi.org/10.1007/978-3-642-41924-9_27.
32. Guizzardi, G., Falbo, R. de A., Guizzardi, R.: Grounding Software Domain Ontologies in the Unified Foundational Ontology (UFO): The case of the ODE Software Process Ontology. CIbSE. (2008).
33. Verdonck, Gailly, F., Poels, G.: 3D vs. 4D Ontologies in Enterprise Modeling. In: Advances in Conceptual Modeling. pp. 13–22 (2014).
34. Verdonck, Gailly, F.: An Exploratory Analysis on the Comprehension of 3D and 4D Ontology-Driven Conceptual Models. In: Conceptual Modeling - ER 2016, Lecture Notes in Computer Science, vol. 9975. pp. 163–172 (2016). https://doi.org/10.1007/978-3-642-33999-8.
35. Mendling, J., Strembeck, M., Recker, J.: Factors of process model comprehension-Findings from a series of experiments. Decision Support Systems. 53, 195–206 (2012). https://doi.org/10.1016/j.dss.2011.12.013.
36. Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., Wesslen, A.: Experimentation in software engineering. (2012). https://doi.org/10.1007/978-3-642-29044-2.
37. Bera, P.: Analyzing the Cognitive Difficulties for Developing and Using UML Class Diagrams for Domain Understanding. Journal of Database Management. 23, 1–29 (2012). https://doi.org/10.4018/jdm.2012070101.
38. Burton-Jones, A., Meso, P.P.N.: Conceptualizing systems for understanding: An empirical test of decomposition principles in object-oriented analysis. Information Systems Research. 17, 38–60 (2006). https://doi.org/10.1287/isre.1050.0079.
39. Voluceau, D. De, Chesnay, L.: Quantitative Measurements of the In ¯ uence of Participant Roles during Peer Review Meetings. Main. 143–159 (2001).
40. Poli, R., Healy, M., Kameas, A.: Theory and applications of ontology: Computer applications. Springer (2010).
41. Herre, H., Loebe, F.: A meta-ontological architecture for foundational ontologies. Lecture Notes in Computer Science. 3761 LNCS, 1398–1415 (2005). https://doi.org/10.1007/11575801_29.
42. Verdonck, M., Prince Sales, T., Gailly, F.: A Comparative Illustration of Foundational Ontologies: BORO and UFO. In: 10th International Conference on Formal Ontology in Information Systems (FOIS 2018). CEUR-WS. org (2018).
43. Sider, T., others: Four-dimensionalism: An ontology of persistence and time. Oxford University Press on Demand (2001).
44. Pease, A., Niles, I.: IEEE standard upper ontology : a progress report. The Knowledge Engineering Review. 17, 65–70 (2002).
45. Allen, J.F.: Maintaining knowledge about temporal intervals. In: Readings in qualitative reasoning about physical systems. pp. 361–372. Elsevier (1990).
46. Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., Wesslen, A.: Experimentation in software engineering. (2012). https://doi.org/10.1007/978-3-642-29044-2.
47. Verdonck, M., Gailly, F.: Design of an empirical study to measure the comprehension of 3 D and 4 D ontology-driven conceptual models. Presented at the International workshop on Value modeling and business ontologies (2017).

48. Moody, D.L.: The Method Evaluation Model : A Theoretical Model for Validating Information Systems Design Methods. Information Systems Journal. 1327–1336 (2003).
49. Vessey, I., Conger, S. a.: Requirements specification: learning object, process, and data methodologies. Communications of the ACM. 37, 102–113 (1994). https://doi.org/10.1145/175290.175305.
50. Burkhardt, J.M., Détienne, F., Wiedenbeck, S.: Object-oriented program comprehension: Effect of expertise, task and phase. Empirical Software Engineering. 7, 115–156 (2002). https://doi.org/10.1023/A:1015297914742.
51. Maes, A., Poels, G.: Evaluating quality of conceptual modelling scripts based on user perceptions. Data & Knowledge Engineering. 63, 701–724 (2007). https://doi.org/10.1016/j.datak.2007.04.008.
52. Evermann, J., Wand, Y.: Ontological Modeling Rules for Uml: an Empirical Assessment. Journal of Computer Information Systems. 47, 14–29 (2006).
53. Falessi, D., Juristo, N., Wohlin, C., Turhan, B., Münch, J., Jedlitschka, A., Oivo, M.: Empirical software engineering experts on the use of students and professionals in experiments. Empirical Software Engineering. 1–38 (2017). https://doi.org/10.1007/s10664-017-9523-3.
54. Shao, J., Wang, H., Chow, S.-C.: Sample size calculations in clinical research. Chapman & Hall/CRC (2008).
55. McKnight, P.E., Najab, J.: Mann-Whitney U Test. Corsini Encyclopedia of Psychology. (2010).
56. Shanks, G.G., Tansley, E., Nuredini, J., Tobin, D., Weber, R.: Representing Part-Whole Relations in Conceptual Modeling: An Emperical Evaluation. MIS Quarterly. 32, 553–573 (2008).
57. Krieger, H., Kiefer, B., Declerck, T.: A Framework for Temporal Representation and Reasoning in Business Intelligence Applications ∗. In: AAAI Spring Symposium: AI Meets Business Rules and Process Management. pp. 59–70 (2008).