Journal of Database Management
Volume 31 • Issue 4 • October-December 2020

A Meta-Analysis of Ontological Guidance and Users' Understanding of Conceptual Models

Arash Saghafi, Tilburg University, The Netherlands
Yair Wand, University of British Columbia, Canada

ABSTRACT

Information systems are intended to be faithful accounts of real-world applications. As an integral part of the development process, analysts create conceptual models in order to understand the application and communicate requirements. Failure to do so has been a prominent reason for IT projects' failure. Hence, improving the quality of models could have a major impact on the information systems' success. To guide the modeling process, researchers use ontology to create more expressive representations of reality. However, improving expressiveness can make the models complicated and cause cognitive hurdles for users. Therefore, the question is whether ontological guidance is worth the trade-off between *expressiveness* and *complexity*. This paper describes a meta-analysis of empirical research examining the impact of ontological guidance on users' understandability. The results show that ontological guidance can improve users' understanding of conceptual models, especially those requiring deeper understanding, thus providing support for ontological guidance in conceptual modeling.

KEYWORDS

Cognition, Conceptual Modeling, Design, Meta-Analysis, Ontology, Representation Theory, Systems Analysis, User Understanding

INTRODUCTION

Systems analysts create conceptual models in order to understand information system (IS) application domains and communicate system requirements (Mylopoulos, 1992) with stakeholders, analysts, designers, and implementers. Failing to understand the domain requirements is a major cause of failure in IS development projects (Wand & Weber, 2002, p. 363). Correcting an error in understanding user requirements post-implementation of the IS is "100 times more costly than it is to correct it during requirements analysis" (Moody, 2005, p. 245). Thus, by enhancing the quality of conceptual models and one could expect a major impact on the success of IS projects.

Conceptual models are required to provide a faithful representation of the relevant aspects of the domain (Wand & Weber, 2002). Use of ontology, "a branch of philosophy that deals with the order and structure of reality in the broadest sense possible" (Angeles, 1981), has been proposed to guide what ought to be modeled (Wand & Weber, 1989) since they "account for the structure and behavior of the world in general" (Storey, 2017, p. 19). Models that are ontologically valid are considered to be more faithful to the reality and thus more 'ontologically expressive'[1] (Wand & Weber, 1993). However, a more expressive grammar tends to have additional constructs (in order to provide a more complete mapping between the grammar and constructs in the ontology) as well as guidelines to

DOI: 10.4018/JDM.2020100103

make sure that the created model is clear (Wand & Weber 1993, p. 228). Thus, Wand and Weber (1993) posited that "the goals of expressive adequacy and simplicity are often in conflict. Additional constructs and production rules enhance expressive power at the cost of increased complexity" (p. 234). The increased complexity might interfere with the process of creating conceptual models, and in interpretation of the models by users. Overall, modelers will face a trade-off between *expressiveness* and *simplicity/parsimony* (Khatri, Vessey, Ram & Ramesh 2006; Bowen, O'Farrell, & Rohde 2009[2]).

The objective of the current paper is to investigate whether conceptual models that are more ontologically expressive can lead to better user understanding despite the possibility of adding to complexity. We use the term 'ontological guidance' to refer to conceptual models where the creators of the model sought guidance from ontology and tried to create conceptual models that are more faithful to reality (i.e., more ontologically expressive). Evaluating the clarity and completeness aspects of a representation (e.g., a conceptual model) is not contingent on using a particular ontological theory.

Systematic investigation of the value of using ontology requires that empirical work to have been previously conducted. Practically, all empirical work has been done using Bunge ontology (Bunge, 1977) (as adapted to information systems by Wand and Weber (1989, 1993, 1995, 2002)). Bunge's ontology is considered the most widely used ontology in systems analysis and design and in conceptual modeling research (Allen & March, 2006a; Fonseca, 2007). Besides the wide adoption of this ontological theory in the IS discipline, Weber (1997) as well as Tilakaratna and Rajapakse (2017) consider Bunge's ontology to be the most complete and best formulated ontology to evaluate information systems analysis and design (Weber, 1997 p. 33; Tilakaratna & Rajapakse, 2017, p. 2). The present paper synthesizes past empirical work without making any claims regarding merits of different ontological theories.

A large body of work has focused on development of ontological guidelines for different conceptual modeling grammars (e.g., Bera, Burton-Jones, & Wand 2011; Evermann & Wand 2006, Recker, Indulska, & Green 2006 – see Table 1) as well as evaluation of the effectiveness of the guidelines on users' performance of cognitive tasks using conceptual models. An important issue has been the study of the trade-off between simplicity and expressiveness. For example, Bodart, Patel, Sim, and Weber (2001) evaluated the ontological expressiveness of entity-relationship (ER) diagrams using different measures of understanding. They showed that for certain measures (e.g., recall) simpler models led to better performance by subjects, while for problem solving tasks, the more expressive models were advantageous. Bowen, et al. (2009) investigated the effect of ontological guidance on models with varying complexities and found that users who write queries on larger but non-ontologically guided models tend to outperform ones using guided models.

This paper investigates the usefulness of applying ontological guidance (based on Bunge's ontology) by conducting a quantitative and objective review of previous empirical work. The work we used had evaluated the effect of ontological guidance on users' performance in order to better understand the trade-off between simplicity and expressiveness. We synthesized papers with different scopes (in terms of modeling languages). They all had the same overall theme of evaluating the impact of ontological guidance. We performed a meta-analysis on previous empirical work[3] (Borenstein, Hedges, Higgins, & Rothstein 2011) and we used the random-effects model.

In an earlier paper, the authors (Saghafi & Wand, 2014) gathered a small number of empirical studies about the impact of ontological guidance on users' understanding of conceptual models. There are both theoretical and empirical differences between the current paper and the earlier version, which lead to differences in scope, method, and depth of analysis. In the previous version we analyzed eight papers. Addressing the limited number of papers in the previous meta-analysis, we extended it by including 10 more papers to a pool of 18 studies, where many studies reported multiple effect sizes, leading to a total of 58 reported effect sizes. We also used a different meta-analysis model - the random effects model – discussed in Choice of Analysis Model section. We then statistically synthesized the reported effect sizes and presented the results using an elaborate categorization of dependent variables based on similarity of measures. We have now four theoretical levels of analysis.

We acknowledge the diversity of the papers included in the meta-analysis (from different modeling grammars to varying dependent variables measured). We incorporated diverse studies in our meta-analysis, as discussed by Rosenthal and DiMatteo (2001), which might be analogized as mixing apples and oranges. Borenstein et al. (2011) state that "combining apples and oranges makes sense if your goal is to produce a fruit salad. The goal of a meta-analysis is only rarely to synthesize data from a set of identical studies" (Borenstein et al. 2011, p. 357).

Building on this argument, this diverse range of studies in our meta-analysis allows us to synthesize empirical studies on the subject of ontological guidance in conceptual modeling. By doing so, we hope to achieve an overarching view of the research and impact of ontological thinking in conceptual modeling.

Using meta-analysis and our categorization of dependent variables, we studied situations where the effect of ontological guidance could vary depending on the cognitive requirements of the task. Our analysis shows that ontological guidance, although may add cognitive complexity, can actually improve users' understanding of conceptual models in tasks that require deeper understanding of the domain.

In the following sections, we first provide a brief review of different ontological foundations used in the information systems literature. We used Bunge's ontology because of its available empirical work. Then, we study the effect of ontological guidance on different aspects of users' performance by synthesizing the prior empirical research. Our findings address the trade-off between simplicity and complexity of conceptual models. Finally, we summarize the results and point to possible conclusions.

APPLICATION OF ONTOLOGY IN THE IS DISCIPLINE

Philosophers have been studying the question of what exists in the world since ancient times (Almeida 2013), from Aristotle to contemporary philosophers (e.g., Bunge 1977; Searle 2006). The branch of philosophy that deals with the structure of reality in the broadest sense is called ontology (Angeles 1981). Here we focus only on applications of ontology in information systems. The underlying premise is that information systems are representations or models of real-world applications (Wand & Weber 1993). Thus, success of an information system is contingent on how effectively and faithfully the representations are generated and interpreted by analysts and designers (Moody 2005; Wand & Weber 1995). In order to build faithful representations of reality, modelers can seek guidance from ontologies (Allen & March, 2006a; Fonseca, 2007) to base their models on the assumed structure of reality.

In Bunge's ontology, we consider the world to be made of things that possess properties. The properties are represented as attributes by human observers. The values of those attributes constitute the state of the thing at a given time (Wand & Weber 1990). The combinations of attribute-values that are possible within a domain are called lawful states. An event describes changes in the state of a thing (Wand & Weber 1993), and are subject to their own laws. An information system is also a thing (representing the real-world) where "lawful states of the information system should reflect the lawful states of the real-world system" (Wand & Wang 1996, p. 89).

IS scholars have used other ontologies as well. Most notable are Allen and March (2006a, 2006b, 2012), who are proponents of using Searle's ontology (Searle 2006) instead of Bunge's. The focus of Searle's ontology (Searle 2006) is on social or institutional reality. The institutional reality is the "world of conceptual objects created by human intentionality and the characteristics ascribed to material or conceptual objects for human purposes" (Allen & March 2006a, p. 1). Allen and March (2006a) believe that "the domain of Bunge's ontology is the physical world, and it has no place for human intentions, interpretations, or meaning" (p. 1). Therefore, it is unfit to present the social concepts[4]. Despite this claim, Volume 4 of Bunge's *Treatise on Basic Philosophy* (Bunge, 1979) is about systems, which also include social systems. In Chapter 5, Bunge examines "the social aspect of man, centering [his] attention on social relations and the resulting social structure" (p. 187)[5].

Use of another ontological theory is discussed by Guizzardi, Herre, and Wagner (2002). They used the general ontological language (GOL) proposed by Degen, Heller, Herre, and Smith (2001). GOL is based on *set theory* and divides the "world into two sorts of entities. On the one hand are urelements, which form an ultimate layer of entities lacking any set-theoretic structure in their make-up. On the other hand are sets, which rise above these urelements in the familiar cumulative hierarchy" (Degen et al., 2001, p. 35). Urelements are "individuals" with properties such as quantity, space, and time. Sets are constructs that are made of individuals - akin to the "kind" concept in Bunge's ontology. GOL is similar to Bunge's ontology in considering entities as building blocks of the world.

To demonstrate the similarities, Guizzardi, et al. (2002) mapped constructs of GOL to Bunge and identified counterparts in respective ontologies. Later, Guizzardi, Wagner, and Sinderen (2004) proposed that ontological guidelines based on GOL could be used in the Unified Modeling Language (UML) grammar. Similarly, Evermann and Wand (2006) used Bunge's ontology to propose ontological guidelines in UML. To compare these guidelines (based on GOL and Bunge's ontology), Hadar and Soffer (2006) performed a qualitative analysis of the UML class diagrams created by 11 software developer professionals. They identified seven different types of modeling variations (six related to syntax and semantics and one different naming conventions)[6] with vagueness in the guidelines "for deciding how to map reality into modeling constructs" (Hadar & Soffer 2006, p. 568). They examined how ontological guidelines can reconcile (or prevent) these variations. The framework by Evermann and Wand (2006) had conclusive rules for five of those variation types and implicit guidance for one. The guidelines developed by Guizzardi et al. (2004) provided conclusive rules for two types of variations, partial rules for another two, implicit guidance for one, and no guidance for the remaining two variation types. Based on Hadar and Soffer's analysis, guidelines rooted in Bunge's ontology have a wider coverage and applicability in reconciling modeling variations.

In addition to the debates on types of ontologies to be used, there are empirical works that demonstrate that employing ontological guidance could make the models complicated and thus negatively affect users' performance. Bodart et al. (2001) showed that users had lower recall of ontologically guided models in comparison with non-guided models because ontologically guided models tend to have more constructs present. Using data analysis, Bowen et al. (2009) studied users' ability to formulate data queries using conceptual models. They moderated the complexity of their tasks by presenting simple models (with fewer constructs) and complex models (with higher number of constructs) to their subjects. Their experiment showed that ontological guidance could improve the accuracy of users' queries on databases (i.e., correctness of the result) based on simple models, while it had a negative effect on performance of users working with more complex models. They showed that ontological guidance could overcomplicate data models and might negatively affect users' queries' accuracy.

Users' ability to formulate queries based on logical data models is not necessarily equivalent to users understanding of conceptual models (Bowen et al. 2009; Tilakaratna & Rajapakse, 2017). Database query studies (Allen & March 2006; Bowen, O'Farrell, and Rohde 2004; 2009) had evaluated the impact of using ontological guidance on users' ability to use "logical database models" (i.e., query databases). Database models are representations of how data is modeled in a database management system, and are used for formulating database queries. Schema of a normalized database may not be necessarily easier to understand by a business analyst (who is not necessarily a database administrator). Moreover, as Bowen et al. 2009 point out, "the primary external factor affecting performance [when using logical database models] is the participants' knowledge of SQL with minimal impacts related to creativity or business experience" (p. 575). Queries may be run on a logical database model that is semantically void (similar to Parsons' 2011 study that showed relationships between entities named 'alpha' and 'beta'). However, a deep-level problem solving question about a domain cannot simply be answered without integrating information from the model with prior knowledge and experiences. In short, these types of models are not within the scope of our paper, as the primary purpose is not necessarily human users' understanding of the model.

Motivation

Our meta-analysis intends to study the trade-off of using ontological guidance where conceptual models becoming *more faithful* to reality and at the same time *more complex*. As mentioned, we use Bunge's ontology because it is the only ontology to have substantial empirical work to test (Weber, 1997, Wand and Weber 2017).

Prior research focused on the development of ontological guidelines (i.e., production rules) based on Bunge's ontology for different conceptual modeling grammars (e.g., Recker et al. 2006, Evermann & Wand 2006, and Bera et al. 2011; see Table 1). The approach has also been used to evaluate the effectiveness of the guidelines on users' performance of cognitive tasks using conceptual models (Bodart et al., 2001; Burton-Jones & Meso, 2006). Ontological guidance has been manifested in two forms. First and foremost being the idea of "ontological expressiveness", that an ontologically clear and complete conceptual modeling grammar can generate better modeling scripts (Wand & Weber, 1993). Ontological expressiveness has been used to guide conceptual modelers when using grammars such as ER (Bodart et al., 2001; Bowen et al. 2004), UML (Evermann & Wand 2006), and business process models (Recker, Rosemann, Green, & Indulska 2011). Second, ontological guidance was applied to good decomposition principles (Wand & Weber 1990) and evaluated by Burton-Jones and Meso (2006, 2008). Decomposition means breaking down the system structure to levels of detail intended to help understand phenomena of interest within the system (Wand & Weber 1995).

While we used papers related to BPMN, ER, OWL, UML and DFDs, they all had the same overall theme of evaluating the impact of ontological guidance on understanding the resulting model. Ontological guidance was used as an independent variable in the random-effects meta-analysis model. We note that the theoretical principles behind ontological expressiveness and good decomposition model do not depend on the Bunge–Wand–Weber (BWW) ontology and could be applied to evaluate the expressiveness of models referencing other ontologies (e.g., GOL).

Introducing the Papers Gathered for the Meta-Analysis

Papers included in this meta-analysis investigated various aspects of the application of ontological guidance in IS modeling. All the papers in our pool focused on the task of interpreting the models – as opposed to creating conceptual models – for the purpose of understanding the application domain (Mylopoulos 1992). The task of understanding may have different requirements; some may be performed by just using the material from the model, while some may require integration of users' prior knowledge with the given material in order to perform the experimental task. These will be further discussed in the Cognitive Theories and the Selection of Papers in the Pool sections.

As shown in Table 1, some studies focused on ontological clarity, some on completeness, and some on both dimensions of ontological expressiveness (Wand & Weber, 1993[1]). The authors only reported independent and dependent variables that were directly related to application of ontology and users' understanding of conceptual models. Some of the papers in the pool had included other variables as covariates or moderators in their hypotheses testing. Bodart et al. (2001), Gemino and Wand (2005) and Bera, Burton-Jones, and Wand (2014) had considered task complexity as a covariate. Khatri et al. (2006) and Burton-Jones and Meso (2008) had investigated the role of prior domain knowledge on users' performance. We did not incorporate these covariates into the meta-analysis since only few papers had measured them.

The dependent variables focused on different aspects of users' understanding of conceptual models. Table 2 presents a list of various dependent variables used by the studies in the meta-analysis pool along with the definition of each measure.

Dependent variables used in these studies focus on different aspects of understanding (e.g., perceptions of understanding and knowledge identification – terms used in the source material). Thus, we categorized the measures based on similarity of scope in order to analyze them from a higher level of abstraction. As shown in Figure 1, the highest level of abstraction aggregated all variables that reflected users' understanding of conceptual models. The decomposition of variables in lower levels is based on different dimensions of cognition that are discussed in some of the cognitive theories used by the papers in our pool (see the Cognitive Theories section).

For the second level in this abstraction, the variables that objectively measured users' understanding of conceptual models (by evaluating users' answers to questions) were placed in the category of actual understanding, while the variables that relied on users' self-reported subjective evaluation (e.g., perceived understanding or confidence in correctness of answers) were placed in the category of perceived understanding.

The third level of this hierarchy focuses on different types of actual understanding of conceptual models. Using the distinction made by Mayer (2003), the tasks that could be performed based solely on the presented material were categorized as surface-level understanding. The tasks that required integration of prior knowledge with information presented in the experimental material were classified as deep-level understanding (manifested in variables such as problem solving, knowledge identification, and quality evaluation effectiveness, as defined in Table 2). The papers in our meta-analysis pool did not make the distinction between perceptions of surface-level understanding and perceptions of deep-level understanding. Hence, the authors do not use surface vs. deep level as a grouping factor for perceptions of understanding.

Finally, the fourth level of the hierarchy categorizes variables that measure performance of tasks that can be done by referring solely to the material from the presented models (i.e., surface-level understanding). The authors considered three categories at this level: First, recall accuracy; this variable, as used by Bodart et al. (2001), measures the number of constructs from the model (e.g., entities, relationships, cardinalities) that participants can recall from memory after the conceptual model is taken from them. The other two categories at this level distinguish between objective evaluations of tasks that can be done using the presented model. One category encompasses experimental questions that were related to some aspects of the domain – these variables were categorized as surface-level *understanding of the domain* (as appearing in the model itself). The other relates to variables that can be done without relying on the semantics of the domain (e.g., models that were void of semantics (Parsons & Cole 2005)). We named those *surface level understanding of the model*.

In short, although the dependent and independent variables and the experiments in these papers were not identical, they were focusing on one abstract question: would presence of ontological guidance improve users' understandability of conceptual models? The main purpose was to see if ontological guidance is 'worth' the added cognitive effort by the users in interpreting them. This motivated us to synthesize the findings of past empirical works on this topic and analyze their findings.

Cognitive Theories

The ontological expressiveness theory discusses levels of completeness and clarity of models. Since the dependent variables (from studies in the meta-analysis pool) are all related to some aspect of cognition, some researchers use cognitive theories to justify their hypotheses regarding users' understanding of conceptual models. Below, we discuss theories of multimedia learning, semantic network, and cognitive fit. We conclude the section by making a comparison of predictions made by the three theories.

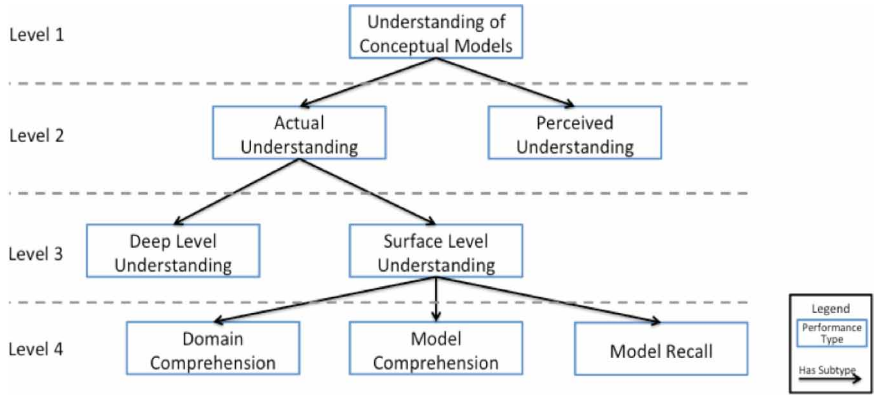
Cognitive Theory of Multimedia Learning (CTML)

Mayer's (2003) theory of multimedia learning (from the field of educational psychology) has been used frequently to evaluate conceptual modeling grammars. The underlying premise is that "when conceptual models include both words and graphic elements, they can be considered multimedia messages" (Gemino & Wand 2005, p. 308). This theory makes two propositions: (1) processing of information in the human mind is done through visual (eyes) and auditory (ears) channels and (2) the human cognitive capacity is limited (Mayer 2003). The theory suggests that learning is achieved when the presented material is integrated with previous knowledge (Mayer 2003).

Table 1. Studies included in the meta-analysis

Authors	Independent Variable(s)	Nature of Study	Task	Dependent Variable(s)	Cognitive Foundations
Bera et al. (2011)	Ontological guidance on OWL ontologies	Intra-grammar – OWL ontologies	Interpreting the model	- Knowledge identification - Perceived understanding - Perceived ease of understanding	Cognitive fit, Multimedia learning
Bera et al. (2014)	- Ontology-based rules - Domain knowledge	Intra-grammar – ER	Interpreting the model	- Problem solving	Multimedia learning
Bodart et al. (2001)	Optional vs. Mandatory properties	Intra-grammar – ER	Interpreting the model	- Recall accuracy - Response accuracy - Response time - Problem solving	Semantic networks, Multimedia learning
Burton-Jones and Weber (1999)	- Ontological clarity - Domain knowledge	Intra-grammar – ER	Interpreting the model	- Problem solving - Perceived ease of understanding	Problem solving
Burton-Jones and Weber (2003)	Ontological clarity	Intra-grammar – ER	Interpreting the model	- Domain comprehension - Confidence	None
Burton-Jones and Meso (2006)	Good Decomposition Model	Intra-grammar – ER	Interpreting the model	- Problem solving - Cloze test (domain comprehension) - Perceived ease of understanding	Semantic networks, Problem solving
Burton-Jones and Meso (2008)	Good Decomposition Model	Intra-grammar – ER	Interpreting the model	- Domain comprehension - Problem solving - Perceived ease of understanding	Cognitive fit, Multimedia learning, Cognitive economy
Burton-Jones, et al. (2012)	Optional vs. Mandatory properties	Intra-grammar – UML	Interpreting the model	- Domain comprehension	Cognitive complexity
Evermann and Wand (2006)	Ontological guidance in UML	Intra-grammar – UML	Interpreting the model	- Problem solving - Domain comprehension	Semantic networks, Multimedia learning, Problem solving theory
Gemino and Wand (2005)	- Optional vs. Mandatory properties - Task complexity	Intra-grammar – ER	Interpreting the model	- Problem solving - Cloze test - Perceived ease of understanding	Theory of multimedia learning
Khatri et al. (2006)	Ontological completeness	Intra-grammar – ER	Interpreting the model	- Problem solving - Model comprehension - Perceived ease of understanding - Completion time	Cognitive fit
Milton, Rajapakse, and Weber (2012)	Ontological clarity	Intra-grammar – UML	Interpreting the model	- Quality evaluation	None
Moody (2002a)	Ontological clarity	Intra-grammar – DFD	Interpreting the model	- Model Comprehension - Verification Accuracy - Completion time	None
Moody (2002b)	Ontological clarity	Intra-grammar – DFD	Interpreting the model	- Model comprehension - Documentation Correctness - Perceived understanding - Completion time	None
Parsons (2011)	- Ontological deficiency - Model Semantics - Task complexity	Intra-grammar – ER	Interpreting the model	- Cloze test - Model comprehension - Confidence	Theory of multimedia learning
Recker et al. (2011)	Ontological clarity	Intra-grammar – BPMN	Interpreting the model	- Perceived ease of understanding - Perceived understanding	None
Shanks, Nuredini, Tobin, Moody, and Weber (2002)	Ontological clarity	Intra-grammar – ER	Interpreting the model	- Problem solving - Domain comprehension - Completion time	None
Shanks, Tansley, Nuredini, Tobin, and Weber (2008)	Modeling composites as entities or relationships	Intra-grammar – ER	Interpreting the model	- Problem solving - Completion time - Perceived ease of understanding	None

Figure 1. Performance measures observed in the meta-analysis



Some of the papers in our meta-analysis (e.g., Bodart et al. 2001 and Burton-Jones & Meso 2006) used the distinction made by Mayer (2003) between "surface-level" understanding and "deep-level" understanding. Surface-level understanding refers to tasks that can be performed using the presented material that is retained in the working memory. Deep-level understanding, in addition to the information from the presented material, requires integration of the cognitive model formed in the working memory with prior knowledge in the long-term memory.

CTML measures learning through two variables:

1. Retention or comprehension refers to the ability to use visual and verbal models in the working memory. Gemino and Wand (2005) suggest assessing retention of domain information by asking questions answerable directly from the presented material. They further suggest that since human cognitive capacity is limited, even a simpler model might be too much to retain in working memory and hence, may not be significantly advantageous over a more complex model (as discussed above). They predicted that the more expressive model could produce better verbal and visual models in the working memory and thus could improve domain comprehension by users;
2. Transfer or problem solving refers to the ability to use knowledge gained from the material to solve problems that are related but not directly answerable from the presented material (Gemino & Wand 2005). Problem solving is based on integration of verbal and visual models with prior knowledge in long-term memory. Bodart et al. (2001) and Gemino and Wand (2005) posited that a clearer and more complete model would be better integrated with prior knowledge. This led to the prediction that subjects receiving ontologically guided models would perform deep-level understanding, or problem solving tasks better than subjects who use models not constructed with ontological guidance.

Theory of Semantic Networks

The theory of semantic networks (Collins & Quillian 1969) posits that the "human semantic memory is structured as a network of nodes linked via directed pathways" (Bodart et al., 2001, p. 388). In Bodart et al. (2001), nodes could refer to entities, attributes of entities, classes, or attributes of classes, and the paths could be any type of relationship between nodes. This theory was also used for making predictions by Burton-Jones and Meso (2006) and Evermann and Wand (2006).

Based on the nodes and links conceptualization of semantic networks, Anderson and Pirolli (1984) presented the theory of spreading activation to describe the human cognition process where information is "spread from node to node along network lines" (Anderson & Pirolli 1984, p. 791)

Table 2. Dependent variables in the meta-analysis

Dependent Variable	Explanation	Grouping in Meta-Analysis
Cloze test, and Domain Comprehension	Questions about the domain that are directly answerable from the model (Gemino & Wand, 2005).	Actual performance: Surface level / *Domain*
Model Comprehension	Evaluating what is directly observable from the model. This evaluation is also used for semantically void (Parsons 2011).	Actual performance
Confidence in Correctness of Answers	Subjects' prediction of the correctness of their answers. It can be measured using a Likert scale (such as Allen and March (2002b), Bowen et al. (2009), and Parsons (2011)).	Perceived Cognitive Performance
Quality Evaluation Effectiveness	Correct identification of (all the) defects in a model (Milton et al. 2012, p. 34). This task requires reference to prior knowledge in order to distinguish correct versus defective models.	Actual Performance – Deep Level
Verification Accuracy	"The ability to identify discrepancies between a data model and a given set of user requirements" (Moody, 2002a).	Actual Performance – Deep Level
Model Recall Accuracy	Proportion of conceptual model constructs (e.g. entities, relationships) that participants recalled correctly, divided by the total number of constructs in the presented model (Bodart et al., 2001). This variable was measured after the diagrams were removed from the participants.	Actual Performance – Surface Level – Recall
Documentation Correctness	The level of completeness of written documentation that was based on a conceptual model (Moody, 2002b).	Actual Performance – Deep Level
Perceived Understanding (model or domain)	The effort required to understand a diagram (Burton-Jones & Meso, 2006).	Perceived Cognitive Performance
Perceived Ease of Understanding (model or domain)	The degree to which the subject "believes that using a particular [conceptual model] would be free of effort" or easy to use (Moody, 2002b, p. 219).	Perceived Cognitive Performance
Problem solving	Answering questions that require integration of prior knowledge with what is observable from the presented material (Gemino & Wand, 2005).	Actual Performance – Deep Level
Knowledge Identification	Knowledge is used by agents to determine the actions required to attain their goals (Bera et al., 2011; Newell, 1982). Knowledge identification is "the task of asking the right questions to determine what actions need to be taken to change the current state of affairs" (Bera et al., 2011, p. 885), based on the conceptual model of the domain.	Actual Performance – Deep Level
Response or Completion Time	Time taken by a participant to answer questions (Bodart et al., 2001).	Task completion time

making knowledge available for processing in the human brain. They claim that as the spreading distance (between the nodes) increases, the strength of processing decays. Using these two related theories (semantic networks and spreading activation) in the context of understanding conceptual models, Bodart et al. (2001) considered three relevant factors for making predictions about the relationship between ontological expressiveness of conceptual models and user understanding:

1. **Number of Constructs:** Increasing the number of constructs in the conceptual model could decrease the likelihood that constructs will be recalled by users as the greater number of nodes in

the human semantic network could decay the spread of activation. Thus, using a more ontologically expressive model decreases the likelihood that users recall the constructs (due to complexity);
2. **Facilitating Elaboration:** Elaboration is the cognitive process of establishing paths between the nodes in the semantic network (Collins & Loftus 1975). Using a more expressive model can lead to a clearer and more complete semantic network in the human brain. A better cognitive model can facilitate identification of alternative paths between nodes and improve the elaboration process. In other words, if traversing along one path failed, another one might be identified;
3. **Inferential Reconstruction:** The ability to infer what is plausible in light of information remembered from the model (Bodart et al., 2001). The theory of semantic networks predicts that using a more expressive model can facilitate inferring routes between nodes that might not be directly connected. This aspect is also related to elaboration process, because better inferential reconstruction can be done when better elaboration is achieved.

In short, the first factor predicts that recalling the more comprehensive model would be more difficult (i.e., surface-level model recall), while the second and third factors predict that the more comprehensive model would lead to creation of a clearer and more complete semantic model in the users' mind, thus improving the performance of tasks that require elaboration and inference (i.e., deep-level understanding).

Problem Solving Theory and the Theory of Cognitive Fit

Newell and Simon (1972) suggest that a person's ability to reason depends on the quality of their mental representation of the domain. They assume that the mental representation is constructed as a "problem space" in the person's memory.

Similar to the theory of problem solving, theory of cognitive fit (Vessey 1991) suggests that when individuals need to solve problems in a domain, their performance will improve when the mental representation of the problem matches the representation of the real-world domain (Shaft & Vessey 2006).

Comparison of Predictions Made by the Three Theories

Here the authors refer to our proposed classification scheme of performance measures in Figure 1 and focus on the leaf-nodes in that diagram. Table 3 compares the predictions made on measures of perceived understanding, deep-level understanding, model recall, and surface-level domain and model comprehension. This comparison (Table 3) reflects the breadth of coverage of these theories, and their similarities and differences in predictions on different dependent variables.

META-ANALYSIS METHOD

Choice of Analysis Model

A meta-analysis could be conceptualized as either a fixed- or random-effects model (Borenstein et al., 2011). A fixed-effects model assumes that all the studies in the meta-analysis are identical and they share a common effect size. Any variation that exists between the findings of different studies in the pool would be due to sampling error. "Put another way, all factors that could influence the effect size are the same in all the studies" (Borenstein et al., 2011, p. 63). The random-effects model, on the other hand, incorporates a group of studies in meta-analysis, assuming that they have "enough in common that it makes sense to synthesize the information, but there is generally no reason to assume that they are identical" (Borenstein et al., 2011, p. 69). The variation between different studies is attributed to sampling error as well as to the random effects variable. Random effects variable accounts for the variation between studies, such as the chosen variables for the study, and the experimental methods.

Table 3. Comparing predictions made on measures of understanding

Measure	Cognitive Theory of Multimedia Learning	Semantic Networks Theory	Problem Solving and Cognitive Fit Theories
Perceived Understanding	Bera et al. (2011) used CTML to predict that users would perceive ontologically expressive models easier to understand as there is "less need to mentally reorganize the information to perform the task", since more expressive models are better domain representations.	No predictions made using this theory in our pool of papers.	Burton-Jones and Meso (2008) posited that ontologically expressive models would provide a more expressive representation of the problem space; thus, users would "perceive the [problem solving] effort to be worthwhile" (p. 754) compared to less expressive models.
Deep-level Understanding	Based on this theory, Gemino and Wand (2005) predicted that ontological expressiveness of a representation facilitates the integration of the model with prior memory. This leads to better problem solving, or in other words, deep level understanding.	Ontologically guided models lead to more precise semantic networks in users' minds. Tasks requiring "a deep-level understanding of a domain mean they must forcefully engage elaborative and inferential reconstruction cognitive processes" (Bodart et al. 2001, p. 389).	Burton-Jones and Meso (2006) claimed that more ontologically expressive models enhance analysts' ability to construct more accurate problem spaces in their memories, which in turn enables them to perform better in problem solving tasks.
Model Recall	No predictions made using this theory in our pool of papers.	According to Bodart et al. (2001), the more ontologically expressive model tends to have "a larger number of construct instances to be remembered, which undermines a user's ability to recall" (p. 389). On the other hand, Burton-Jones and Meso (2006) predicted that recall of more ontologically expressive models would be more accurate since these models enable "faster and more accurate recall" (p. 45) due to the strength of the connections between nodes in the semantic network in a subjects' memory.	Surface-level understanding is not within the scope of this theory.
Surface-level Domain Comprehension	Burton-Jones and Meso (2008) proposed that less ontologically expressive models require extraneous cognitive load by users in order to comprehend the relevant information from the model. Thus, less ontologically expressive models inhibit users' comprehension.	Bodart et al. (2001) claimed that ontologically expressive models would afford better elaboration processes and positively affect users' comprehension.	Surface-level understanding is not within the scope of this theory.
Surface-level Model Comprehension	No predictions made using this theory in our pool of papers, but using this theory, one could make predictions similar to domain comprehension.	No predictions made using this theory in our pool of papers, but using this theory, one could make predictions similar to domain comprehension.	Surface-level understanding is not within the scope of this theory.
Task Completion Time	No predictions made using this theory in our pool.	No predictions made using this theory in our pool.	No predictions made using this theory in our pool.

The studies gathered in our pool had different independent and dependent variables, yet they all focused on the *influence of ontological guidance on some measures of cognitive tasks performed by users*. Because the studies in the pool are not identical, we chose the random-effects model for this meta-analysis. In our initial study (Saghafi & Wand, 2014), the authors chose the fixed-effects model, as recommended by Borenstein et al. (2011), since the sample was considerably smaller.

Selection of Papers in the Pool

Using online databases[7] – namely Business Source Complete, Web of Science, and JSTOR – and data from a paper by Burton-Jones, Green, Indulska, Recker, and Weber (2017), titled "Assessing Representation Theory with a Framework for Pursuing Success and Failure"[8], we ended up with a pool of 314 papers. From this pool, we selected papers where the researches had actually conducted empirical experiments using ontologically guided conceptual models that were rooted in Bunge's ontology. We set the scope of the meta-analysis to empirical studies that focused on model *interpretation* rather than *creation*[9]. Thus, we eliminated papers describing research on users' ability to create models (e.g., Hadar & Soffer 2006). We conducted the meta-analysis with 18 papers (including 58 reported effect sizes) of empirical work on understanding conceptual models[10].

Variables Used in the Study

As illustrated in Figure 1 and Table 2, the studies in the meta-analysis reported different dependent variables. We excluded recall accuracy and task completion time. Recall accuracy[11] was excluded because we assume that analysts in the real world will have access to models during the analysis and design process and rarely need to recall models from memory (Parsons and Cole, 2005). Considering this factor (i.e., constant access to the models), we also excluded a problem solving measure from Bodart et al. (2001), as subjects did not have access to conceptual models for that portion of the study (i.e., Experiment III in Bodart et al. 2011); however, problem solving measures from other papers in our meta-analysis pool were included.

As for task completion time, there were no predictions in the cognitive theories about the relationship between presence of ontological guidance and task completion.

Table 2 provides a list of the dependent variables that were incorporated in our meta-analysis. For this study, we chose a meta-dependent-variable called "*understanding of conceptual models*", which incorporated all the dependent variables from Table 2 (excluding recall accuracy and task completion time as stated above).

Meta-Analysis Hypotheses

Using the predictions of Table 3, we propose four potential hypotheses to test in this meta-analysis. Our first hypothesis is regarding users' perceptions of understanding of conceptual models, which relates to the partitioning of cognitive performance measures (between perceptions of understanding and actual understanding - see Figure 1). We make a similar prediction to the one made by Burton-Jones and Meso (2008) based on the problem solving theory (Newell & Simon 1972). We predict that users of ontologically guided models would have a more expressive representation of the problem space and would "perceive the [problem solving] effort to be worthwhile" (Burton-Jones & Meso, 2008, p. 754) compared with users of less expressive models. Hence:

Hypothesis 1: Users will have higher *perceptions* of understanding when working with ontologically guided models compared with non-guided models.

The second hypothesis focuses on tasks that require deeper levels of understanding of conceptual models. Referring to the CTML (Mayer, 2003), users need to integrate the information from conceptual models with their prior knowledge in order to solve problems. We predict that retention and transfer of model information is facilitated when the models are clearer and more complete with respect to the structure of reality (i.e., more ontologically expressive). Thus:

Hypothesis 2: Users of ontologically guided models will achieve deeper levels of understanding compared with those using non-guided models.

Hypotheses 3 and 4 revolve around measures of surface-level understanding. As discussed earlier, some of the papers in our meta-analysis pool tasked subjects with activities that could be done using only the presented model, without requiring integration with prior knowledge (e.g., Shanks et al. 2002, Burton-Jones & Meso 2006). There were also surface-level understanding tasks that could be completed without relying on the semantics of the domain, which we grouped under surface-level understanding of the models category, or model comprehension (e.g., Moody 2002a; b, Evermann & Wand 2006). This measure is also used when semantically void models are presented to subjects (as done by Parsons (2011)). Although the number of studies is limited, we included this variable in our hypothesis testing since it is an aspect of users' understanding of conceptual models.

Similar to how Bodart et al. (2001) used theory of semantic networks to hypothesize about users' surface-level understanding, we claim that more expressive models can lead to better (or more expressive) cognitive models in subjects' working memory, and thus they would be better able to identify alternative paths between nodes in the semantic network. Hence, a more ontologically expressive model can improve the elaboration process.

Hypothesis 3: Users of ontologically guided models will achieve better comprehension of *domains* compared with users of non-guided models.
Hypothesis 4: Users of ontologically guided models will achieve better comprehension of *models* compared with users of non-guided models.

Data Analysis

As mentioned earlier (Table 1), the papers in our meta-analysis pool had measured different dependent variables in their empirical experiments. Besides the difference in types of variables, the findings were also reported using different statistical measures (e.g. *t*-values, *F*-values, and regressions coefficients). To synthesize the reported measures, we converted all the reported measures to Cohen's *d*[12], which is the standardized mean difference between two experimental groups – in this case ontologically guided models vs. non-guided models. Cohen's d demonstrates the difference between the two group's means in terms of standard deviation (e.g., Cohen's d of 1.00 means that treatment group's mean is one standard deviation greater than the control group's mean). We abstracted all the (included) dependent variables to "understanding of conceptual models" and synthesized the reported effect sizes – as reported in Table 4 and discussed below in more detail. In addition to the first round of analysis (i.e., encompassing all the variable types), we grouped similar dependent variables (the schema in Figure 2) and performed three additional rounds of analysis.

We also point out that except for the experiments done by Burton-Jones and Meso (2006), which included three levels of good, moderate, and bad decomposition (or ontological guidance), the remaining papers in our pool only compared ontologically-guided and non-guided models (i.e., two levels). We are not claiming that non-guided models are completely void of ontology, as ontological thinking about things in the world, their attributes, and events are ingrained in all modeling grammars. However, we treated the specific presence of ontological guidance – which leads to clearer and more complete models – to be a binary variable in our meta-analysis[14].

The first round of meta-analysis used 18 papers, with 58 reported effect sizes in total (Table 4). The Cohen's *d* was 0.56, which indicates that on average, ontological guidance can improve understanding of conceptual models by 0.56 standard deviations (compared with the performance

Table 4. Meta-analyses done in this study

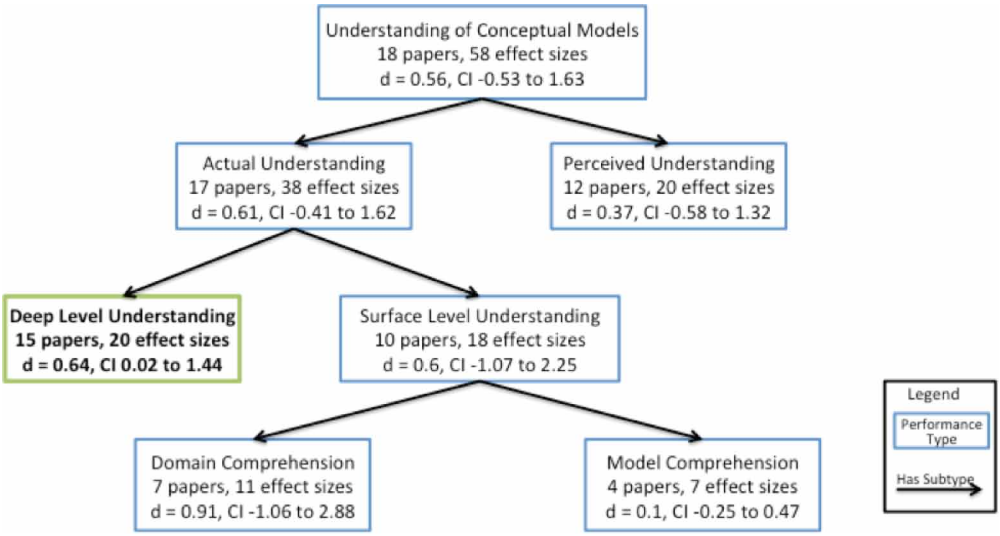
Round	Focus of Analysis (as Named in the Hierarchy in Figure 1)	No. of Papers	No. of Effect Sizes	No. of Dependent Variables	Cumulated Subjects by Effect Size	Average Cohen's *d*	Confidence Interval (95%)
1	Cognitive Performance on Conceptual Models	18	58	9	1648	0.56	–0.53 to 1.63
2	Actual Cognitive Performance	17	38	7	1025	0.61	–0.41 to 1.62
	Perceived Cognitive Performance	12	20	2	1216	0.37	–0.58 to 1.32
3	Actual Performance - Surface Level	10	18	2	603	0.6	–1.07 to 2.25
	Actual Performance - Deep Level[13]	14	20	5	791	0.64	0.02 to 1.44
4	Actual Performance - Surface Level - Domain	7	11	1	502	0.91	–1.06 to 2.88
	Actual Performance - Surface Level - Model	4	7	1	181	0.1	–0.25 to 0.47

of subjects who used non-guided models). For this random effects meta-analysis (which contains reported effect sizes that are non-identical), we report a 95 percent confidence interval while taking into account the sampling errors of the studies in the meta-analysis pool. The confidence interval contains the distribution where 95 percent (as is common in the literature) "of true effects are expected to be found" (Borenstein et al., 2011, p. 350). For the first round of the meta-analysis, the confidence interval of the average effect size (with 95% certainty) was from –0.53 to 1.63. In other words, 95 percent of the time the range of impact of ontological guidance on subjects' performance is from –0.53 to 1.63 standard deviations (compared with average performance without ontological guidance). The wide confidence interval of our analysis (which includes negative values) is likely due to the random effects component that varies between different studies (as incorporated into the random-effect model in the meta-analysis). By not including the random effects variable (i.e., going with the assumption of having identical studies in the pool and using the fixed-effects meta-analysis model), the confidence interval would have been considerably narrower (0.36 to 0.70, i.e., significant effect on cognitive performance). In order to somewhat reduce the randomness in the analysis, we grouped reported effect sizes based on similarity of scope (following the abstraction hierarchy in Figure 1).

The second round separated actual measures of cognitive performance (i.e., problem solving, knowledge identification, quality evaluation, comprehension) from perceived measures (i.e., perceived understanding, perceived ease of understanding, and confidence in answers). In the meta-analysis on measures of actual cognitive performance, 17 papers were represented containing 38 reported effect sizes. Cohen's *d* was 0.61 with a 95 percent confidence interval of –0.41 to 1.62.

The analysis of perceived measures of cognitive performance was performed on 12 papers with 20 reported effect sizes. Cohen's *d* was 0.37 with a 95 percent confidence interval of –0.58 to 1.32; this confidence interval rejects Hypothesis 1 (i.e., increase in users' perceptions of understanding when using ontologically guided models). The lack of support for this hypothesis could be attributed to reasons such as the random effects (i.e., having multiple non-identical studies with too much variation in the meta-analysis pool), or many of the studies measuring perceptions of understanding after tasks that required either surface-level or deep-level understanding, and in certain cases both types of understandings (e.g., study by Gemino & Wand, 2005).

Figure 2. Average effect sizes of the meta-analysis based on performance measures. *d*: Cohen's *d*; CI: 95% confidence interval.



The third round focused on the actual measures of understanding. Following Mayer's (2003) distinction between surface-level and deep-level understandings, two groups were created. Meta-analysis of surface-level understanding was done on nine papers with 18 reported effect sizes. Cohen's *d* was 0.6 with a 95 percent confidence interval of –1.07 to 2.25.

The most reliable effect was observed in the group of measures focusing on deep-level understanding. This analysis included 14 papers, with 20 reported effect sizes. The average Cohen's *d* was 0.64, with the 95 percent confidence interval of 0.02 to 1.44[15]. Both lower and upper bounds of the interval being positive confirms Hypothesis 2, or in other words, it indicates that ontological guidance has a uniform and positive effect on improving deep-level understanding.

The fourth round focused on the measures of surface-level understanding. In the pool of papers used in the meta-analysis, the authors had distinguished between comprehension of conceptual models at the domain level versus the model level. We followed their suggestions and created two groups of domain and model surface-level understanding. The effect of surface-level understanding of domains was strong (0.91), while the confidence interval is wide (from –1.06 to 2.88), hence rejecting Hypothesis 3 regarding improvements in domain comprehension using ontologically guided models. Even though the Cohen's d of 0.91 might indicate a strong effect, the wide confidence interval could be an artifact of the high degree of variance in the random-effects meta-analysis model.

The surface-level understanding of models (or model comprehension) was represented by only four papers with seven effect sizes (Khatri et al. 2006; Moody, 2002a, 2002b; Parsons, 2011). The average effect size was 0.01 with a 95 percent confidence interval of –0.25 to 0.47. This leads to rejection of Hypothesis 4, which had predicted a positive effect on model comprehension by ontologically guided models.

Figure 2 summarizes the meta-analysis' findings based on different measures of performance.

SUMMARY OF FINDINGS, DISCUSSION AND IMPLICATIONS

The following points summarize our findings:

- Ontological guidance has a significant effect on improving users' deep-level understanding (Cohen's $d = 0.64$ with the confidence interval of 0.02 to 1.44);

- Ontological guidance has weak or no effect on users' (actual) surface-level understanding of models (Cohen's $d = 0.1$) with non-significant statistical results.

The significant effect of ontological guidance on deep-level understanding – which, according to Mayer (2003) requires integration of working memory with prior knowledge – was predicted by the cognitive theory of multimedia learning as well as the theory of semantic networks. According to the theory of multimedia learning, the ontologically expressive models lead to formation of higher quality models in the working memory, which will be better integrated with prior knowledge and result in better deep-level understanding. The theory of semantic networks (Collins & Quillian, 1969) together with the theory of spreading activations (Anderson & Pirolli, 1984) also predict that the more expressive model will lead to better elaboration and inferential reconstruction in the minds of users (Bodart et al., 2001). Therefore, it seems that users' problem solving performance (using conceptual models) improves, despite having additional constructs resulting from employing ontological guidance.

When focusing on surface-level understanding, we observe that model comprehension is not affected (Cohen's $d = 0.1$). To explain this phenomenon, we refer to the distinction between evaluations of the syntax, semantics, and pragmatics of conceptual modeling grammars proposed by Burton-Jones et al. (2009). Syntactic evaluation might "involve examining valid ways in which scripts can be created using a grammar or examining alternative ways that individuals form scripts using the grammar" (Burton-Jones, Wand, and Weber 2009, p. 497). Semantic evaluation examines "the meaning of the constructs in the grammar" (Burton-Jones et al., 2009, p. 497) and how the meaning "can be conveyed more clearly and completely" (Bera et al, 2014 p. 1). Pragmatic evaluation of the conceptual domain reflects the context; more specifically the contextual conditions "in which models are more likely to be understood or preferred" (Bera et al., 2014 p. 1). Using this distinction, we contend that surface-level understanding of the model reflects mainly the syntactic evaluation of the constructs. This was done by Parsons (2011), who removed meaning from the constructs, and instead used symbols such as alpha and beta for constructs and relationships. One could argue that syntactic model comprehension might have lesser ties in this case to the real world, thus weakening the effect of ontological expressiveness (namely, completeness and clarity) on the model's representation of the real world[16].

As for the moderate to weak influence of ontological guidance on perceived understanding, we note that the measures of perceived understanding were collected under varied conditions. For example, Recker et al. (2011) conducted a survey to measure perceptions of professionals without asking them to perform any experimental tasks. Burton-Jones and Meso (2006, 2008) measured perceptions of users after they had completed both problem solving and cloze-test (surface-level understanding) tasks. However, surface-level understanding of the domain reflects meaning as well as the context of usage, specifically, the application domain. In this case, the impact of ontological guidance on users' performance shows a rather large effect size of 0.91, although, it was not statistically significant.

With regards to practical implications, we believe that incorporating ontological thinking in training of analysts could improve their domain understanding in requirements engineering and systems analysis phases. More specifically, if practitioners learn to create more ontologically expressive models, other analysts or future users who refer to the documented models will be able to better understand them – particularly for tasks that require deeper levels of understanding.

Limitations, Strengths, and Weaknesses of the Study

As shown in Table 4, the only significant result (with a positive confidence interval) was related to measures of deep-level understanding (represented by 14 papers, and 20 effect sizes in our meta-analysis pool). The possible reason for a wide confidence interval for other dimensions of understanding could be the limited work on those measures. As mentioned earlier, we dealt with a heterogeneous sample of studies (i.e., different grammars, also different independent and dependent variable) in the meta-analysis pool. "In fact, random variation alone can easily lead to large disparities in *P* values, far beyond falling just to either side of the 0.05" (Amrhein, Greenland, & McShane 2019, p. 306). In addition, limited number of studies for certain tasks (e.g., only seven for surface-level domain comprehension) could be a major reason for the wide confidence intervals that we observed.

Publication Bias

In a systematic review such as our meta-analysis, a particular concern to address is publication bias, described as "the studies with statistically significant results are more likely to find their way into the published literature than studies that report results that are not statistically significant" (Borenstein et al., 2011, p. 278). Unpublished studies are metaphorically "archived in the file-drawer". To address this "file-drawer threat", we can calculate fail-safe *N*. This measure estimates the number of studies with insignificant effect sizes (that might be in the file drawer), and which, if incorporated into the meta-analysis, can reduce the overall effect size (Borenstein et al., 2011, p. 284) to a pre-determined criterion effect size. The literature suggests setting the criterion effect size at 0.1 (Orwin, 1983). In other words, fail-safe *N* would be the number of studies with a zero effect size that, if incorporated into the meta-analysis, could reduce the overall effect size (i.e., Cohen's *d*) to the criterion effect size (e.g., $d_c = 0.1$). We calculated the fail-safe N_{fs}, with respect to the number of reported effect sizes included in the meta-analysis (*K*), using the equation below (Rosenthal & DiMatteo 2002):

$$N_{fs} = [K(d - d_c)]/d_c$$

Fail-safe N for the only significant effect of our meta-analysis (i.e., deep-level understanding) was 108, which indicates the strength of the effect. We need to point out that fail-safe N is not the number of studies that could make our interval cross zero (and become a non-significant effect). It implies that if 108 non-significant findings are added to the study, they could bring down the average effect size of the deep-level understanding from 0.64 down to a pre-set criterion effect size of Cohen's d = 0.1 (as the literature suggested that threshold).

CONCLUSION AND FUTURE RESEARCH

Scholars in the information systems domain have debated whether ontological guidance should be employed because the ontological expressiveness of models comes at the price of losing simplicity. Among various ontological theories used as guidance, the Bunge–Wand–Weber ontology (Wand & Weber, 1989) is the most widely used ontological theory in the information systems field (Allen & March, 2006a, Fonseca, 2007). Other theories have also been discussed in the literature, most notably by Allen and March (2006a, 2006b, 2012), who are the proponents of using Searle's ontology (Searle, 2006). However, these theories do not have a substantial amount of empirical work to justify a meta-analysis. We need to emphasize that our goal was not to support a particular ontology. Due to the precedence set by prior research's focus on Bunge's ontology, we synthesized and reported empirical studies that had studied the BWW ontology. As discussed in the literature, "additional constructs and production rules enhance expressive power at the cost of increased complexity" (Wand & Weber 1993, p. 234). Our goal was to better understand the trade-off between model complexity and users' understandability. We found that in deep-levels understanding (e.g., problem solving) tasks, the trade-off of having perhaps more constructs was indeed worth it, as human performance using ontologically guided models was significantly better.

Because each study in our pool focused on a different aspect of ontological guidance, we chose the random-effect model for this meta-analysis. The random effects variable component between the studies leads to wide confidence intervals for the findings (Amrhein et al. 2019) – see Table 4. Identifying the variations in a treatment is one of the reasons for doing a meta-analysis (Borenstein et al. 2011). We noticed the significant effect of ontological guidance on deep-level understanding of conceptual models. For such tasks the confidence interval of this effect was positive 95 percent of the time. Although we are reflecting findings of past studies, as all meta-analyses do, we also addressed the opposing views about value of ontological guidance and the trade-off between ontological expressiveness and complexity.

Our meta-analysis finds conclusive results in favor of ontological guidance on one dimension of understanding (namely deep-level). We hope our meta-analysis leads other researchers to focus on less charted territories where there were fewer effect sizes (e.g., surface-level understanding of models or domains, perceptions). The causal and moderating factors that influence the impact of ontological guidance could be the subjects of future research to find more conclusive results under varying circumstances.

Moreover, studies similar to that of Hadar and Soffer (2006) regarding the coverage of different ontologies (e.g., BWW vs. Searle) for different modeling grammars (e.g., ER, BPMN) can be conducted to help researchers and practitioners in choosing the ontological theory that best suits their requirements.

ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

Allen, G. N., & March, S. T. (2006a). A critical assessment of the Bunge-Wand-Weber ontology for conceptual modeling. *16th Annual Workshop on Information Technologies and Systems (WITS).*

Allen, G. N., & March, S. T. (2006b). The effects of state-based and event-based data representation on user performance in query formulation tasks. *Management Information Systems Quarterly*, *30*(2), 269–290. doi:10.2307/25148731

Allen, G. N., & March, S. T. (2012). A research note on representing part–whole relationships in conceptual modeling. *Management Information Systems Quarterly*, *36*(2), 945–964. doi:10.2307/41703488

Almeida, M. B. (2013). Revisiting ontologies: A necessary clarification. *Journal of the American Society for Information Science and Technology*, *64*(8), 1682–1693. doi:10.1002/asi.22861

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305–307. doi:10.1038/d41586-019-00857-9 PMID:30894741

Anderson, J. R., & Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *10*(4), 791–798. doi:10.1037/0278-7393.10.4.791

Angeles, P. A. (1981). *Dictionary of philosophy*. Barnes and Noble Books.

Bera, P., Burton-Jones, A., & Wand, Y. (2011). Guidelines for designing visual ontologies to support knowledge identification. *Management Information Systems Quarterly*, *35*(4), 883–908. doi:10.2307/41409965

Bera, P., Burton-Jones, A., & Wand, Y. (2014). Research note – How semantics and pragmatics interact in understanding conceptual models. *Information Systems Research*, *25*(2), 401–419. doi:10.1287/isre.2014.0515

Bodart, F., Patel, A., Sim, M., & Weber, R. (2001). Should optional properties be used in conceptual modeling? A theory and three empirical tests. *Information Systems Research*, *12*(4), 384–405. doi:10.1287/isre.12.4.384.9702

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Bowen, P. L., O'Farrell, R. A., & Rohde, F. H. (2004). How does your model grow? An empirical investigation of the effects of ontological clarity and application domain size on query performance. In *Proceedings from the 25th International Conference on Information Systems* (pp. 77-90). Association of Information Systems.

Bowen, P. L., O'Farrell, R. A., & Rohde, F. H. (2009). An empirical investigation of end-user query development: The effects of improved model expressiveness vs. complexity. *Information Systems Research*, *20*(4), 565–584. doi:10.1287/isre.1080.0181

Bunge, M. (1977). *Treatise on basic philosophy: Ontology I: The furniture of the world* (Vol. 1). Springer. doi:10.1007/978-94-010-9924-0

Bunge, M. (1979). *Treatise on basic philosophy: Ontology II: World of systems* (Vol. 4). Springer. doi:10.1007/978-94-009-9392-1

Burton-Jones, A., Clarke, R., Lazarenko, K., & Weber, R. (2012). Is use of optional attributes and associations in conceptual modeling always problematic? Theory and empirical tests. *ICIS 2012 Proceedings.*

Burton-Jones, A., & Meso, P. N. (2006). Conceptualizing systems for understanding: An empirical test of decomposition principles in object-oriented analysis. *Information Systems Research*, *17*(1), 38–60. doi:10.1287/isre.1050.0079

Burton-Jones, A., & Meso, P. N. (2008). The effects of decomposition quality and multiple forms of information on novices' understanding of a domain from a conceptual model. *Journal of the Association for Information Systems*, *9*(12), 1. doi:10.17705/1jais.00179

Burton-Jones, A., Recker, J., Indulska, M., Green, P., & Weber, R. (2017). Assessing Representation Theory with a Framework for Pursuing Success and Failure. *Management Information Systems Quarterly*, *41*(4), 1307–1333. doi:10.25300/MISQ/2017/41.4.13

Burton-Jones, A., Wand, Y., & Weber, R. (2009). Guidelines for empirical evaluations of conceptual modeling grammars. *Journal of the Association for Information Systems*, *10*(6), 1. doi:10.17705/1jais.00201

Burton-Jones, A., & Weber, R. (1999). Understanding relationships with attributes in entity-relationship diagrams. In *Proceedings of the 20th International Conference on Information Systems* (pp. 214-228). Association for Information Systems.

Burton-Jones, A., & Weber, R. (2003). Properties do not have properties: Investigating a questionable conceptual modeling practice. *Proceedings of the Second Annual Symposium on Research in Systems Analysis and Design*.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. doi:10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. doi:10.1016/S0022-5371(69)80069-1

Degen, W., Heller, B., Herre, H., & Smith, B. (2001). GOL: toward an axiomatized upper-level ontology. *Proceedings of the International Conference on Formal Ontology in Information Systems*, 34-46.

Evermann, J., & Wand, Y. (2006). Ontological modeling rules for UML: An empirical assessment. *Journal of Computer Information Systems*, *46*(5), 14–29. doi:10.1080/08874417.2006.11645922

Fonseca, F. (2007). The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology*, *58*(6), 786–793. doi:10.1002/asi.20565

Gemino, A., & Wand, Y. (2004). A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering*, *9*(4), 248–260. doi:10.1007/s00766-004-0204-6

Gemino, A., & Wand, Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering*, *55*(3), 301–326. doi:10.1016/j.datak.2004.12.009

Guizzardi, G., Herre, H., & Wagner, G. (2002). *On the general ontological foundations of conceptual modeling. Conceptual Modeling—ER 2002*. Springer Berlin Heidelberg.

Guizzardi, G., Wagner, G., & Sinderen, M. (2004). A formal theory of conceptual modeling universals. *Proceedings of the Workshop on Philosophy and Informatics (WSPI)*, 1-10.

Hadar, I., & Soffer, P. (2006). Variations in conceptual modeling: Classification and ontological analysis. *Journal of the Association for Information Systems*, *7*(8), 568–592. doi:10.17705/1jais.00096

Khatri, V., Vessey, I., Ram, S., & Ramesh, V. (2006). Cognitive fit between conceptual schemas and internal problem representations: The case of geospatio-temporal conceptual schema comprehension. *IEEE Transactions on Professional Communication*, *49*(2), 109–127. doi:10.1109/TPC.2006.875091

March, S. T., & Allen, G. N. (2014). Toward a social ontology for conceptual modeling. *Communications of the AIS, 34*, Article 70.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, *13*(2), 125–139. doi:10.1016/S0959-4752(02)00016-6

Milton, S. K., Rajapakse, J., & Weber, R. (2012). Ontological clarity, cognitive engagement, and conceptual model quality evaluation: An experimental investigation. *Journal of the Association for Information Systems*, *13*(9), 2. doi:10.17705/1jais.00307

Moody, D. L. (2002a). Complexity effects on end user understanding of data models: An experimental comparison of large data model representation methods. In ECIS 2002 Proceedings (pp. 482-496). Academic Press.

Moody, D. L. (2002b). Comparative evaluation of large data model representation methods: The analyst's perspective. In *Conceptual Modeling—ER 2002 Proceedings of the 21st International Conference on Conceptual Modeling* (pp. 214-231). Springer.

Moody, D. L. (2005). Theoretical and practical issues in evaluating the quality of conceptual models: Current state and future directions. *Data & Knowledge Engineering*, *55*(3), 243–276. doi:10.1016/j.datak.2004.12.005

Mylopoulos, J. (1992). Conceptual modeling and Telos 1. In P. Loucopoulos & R. Zicari (Eds.), *Conceptual Modeling, Databases, and Case: An Integrated View of Information Systems and Development*. John Wiley & Sons.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*, *18*(1), 87–127. doi:10.1016/0004-3702(82)90012-1

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. doi:10.2307/1164923

Parsons, J. (2011). An experimental study of the effects of representing property precedence on the comprehension of conceptual schemas. *Journal of the Association for Information Systems*, *12*(6), 1. doi:10.17705/1jais.00268

Parsons, J., & Cole, L. (2005). What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. *Data & Knowledge Engineering*, *55*(3), 327–342. doi:10.1016/j.datak.2004.12.008

Recker, J., Indulska, M., & Green, P. (2006). How good is BPMN really? Insights from theory and practice. *ECIS 2006 Proceedings*.

Recker, J., Rosemann, M., Green, P. F., & Indulska, M. (2011). Do ontological deficiencies in modeling grammars matter? *Management Information Systems Quarterly*, *35*(1), 57–79. doi:10.2307/23043489

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Review of Psychology*, *52*(1), 59–82. doi:10.1146/annurev.psych.52.1.59 PMID:11148299

Rosenthal, R., & DiMatteo, M. R. (2002). Meta-analysis. In H. Pashler & J. Wixted (Eds.), Stevens' handbook of experimental psychology, Volume 4, Methodology in experimental psychology (3rd ed.). Academic Press.

Saghafi, A., & Wand, Y. (2014, January). Do ontological guidelines improve understandability of conceptual models? a meta-analysis of empirical work. In *2014 47th Hawaii International Conference on System Sciences* (pp. 4609-4618). IEEE.

Searle, J. R. (2006). Social ontology: Some basic principles. *Anthropological Theory*, *6*(1), 12–29. doi:10.1177/1463499606061731

Shaft, T. M., & Vessey, I. (2006). The role of cognitive fit in the relationship between software comprehension and modification. *Management Information Systems Quarterly*, *30*(1), 29–55. doi:10.2307/25148716

Shanks, G., Tansley, E., Nuredini, J., Tobin, D., & Weber, R. (2008). Representing part-whole relations in conceptual modeling: An empirical evaluation. *Management Information Systems Quarterly*, *32*(3), 553–573. doi:10.2307/25148856

Shanks, G., & Weber, R. (2012). The hole in the whole: A response to Allen and March. *Management Information Systems Quarterly*, *36*(3), 965–980. doi:10.2307/41703489

Shanks, G. G., Nuredini, J., Tobin, D., Moody, D. L., & Weber, R. (2002). Representing things and properties in conceptual modeling: an empirical evaluation. *ECIS 2002: Proceedings of the Xth European Conference on Information Systems: Information Systems and the Future of the Digital Economy*, 1775-1785.

Soffer, P., & Hadar, I. (2007). Applying ontology-based rules to conceptual modeling: A reflection on modeling decision making. *European Journal of Information Systems*, *16*(5), 599–611. doi:10.1057/palgrave.ejis.3000683

Storey, V. C. (2017). Conceptual modeling meets domain ontology development: A reconciliation. *Journal of Database Management*, *28*(1), 18–30. doi:10.4018/JDM.2017010102

Tilakaratna, P., & Rajapakse, J. (2017). Evaluation of the ontological completeness and clarity of object-oriented conceptual modeling grammars. *Journal of Database Management*, *28*(2), 1–26. doi:10.4018/JDM.2017040101

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, *47*(4), 1274–1294. doi:10.3758/s13428-014-0527-2 PMID:25361866

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219–240. doi:10.1111/j.1540-5915.1991.tb00344.x

Wand, Y., Storey, V. C., & Weber, R. (1999). An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions on Database Systems*, *24*(4), 494–528. doi:10.1145/331983.331989

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95. doi:10.1145/240455.240479

Wand, Y., & Weber, R. (1989). *An ontological evaluation of systems analysis and design methods. Information system concepts: An in-depth analysis*. Elsevier Science Publishers BV, North-Holland.

Wand, Y., & Weber, R. (1990). An ontological model of an information system. *IEEE Transactions on Software Engineering*, *16*(11), 1282–1292. doi:10.1109/32.60316

Wand, Y., & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, *3*(4), 217–237. doi:10.1111/j.1365-2575.1993.tb00127.x

Wand, Y., & Weber, R. (1995). On the deep structure of information systems. *Information Systems Journal*, *5*(3), 203–223. doi:10.1111/j.1365-2575.1995.tb00108.x

Wand, Y., & Weber, R. (2002). Research commentary: Information systems and conceptual modeling—a research agenda. *Information Systems Research*, *13*(4), 363–376. doi:10.1287/isre.13.4.363.69

Wand, Y., & Weber, R. (2017). Thirty years later: Some reflections on ontological analysis in conceptual modeling. *Journal of Database Management*, *28*(1), 1–17. doi:10.4018/JDM.2017010101

Weber, R. (1997). *Ontological foundations of information systems*. Coopers and Lybrand and the Accounting Association of Australia and New Zealand.

ENDNOTES

1 An ontologically complete grammar is one in which there is a total mapping between constructs of the conceptual modeling grammar and the ontological concepts (such as thing, property, and event). Ontological clarity is described based on three types of deficiencies: (i) construct overload occurs when a construct from the grammar is used to model two or more concepts from ontology, (ii) construct redundancy occurs if two or more constructs map to the same ontological concept, and (iii) excess construct occurs when a construct from the grammar does not map to any ontological concepts.

2 "For two equivalent models, the more expressive model exhibits greater overall complexity than the parsimonious model" (Bowen, O'Farrell, and Rohde 2009, p. 568).

3 Another approach for synthesizing prior work is conducting *narrative reviews* (i.e., qualitative interpretations of different studies). However, narrative reviews are subjective evaluations and become very difficult when there are more than a few studies involved (Borenstein et al. 2011). Meta-analysis, on the other hand, is a statistical analysis that leads to objective results.

4 As an example, March and Allen (2014) refer to the inception of the state of Utah as an institutional fact that Bunge's ontology – according to them – fails to represent. Based on Searle's ontology (Searle, 2006), a speech act (declaration) by the United States government on January 4, 1896, made Utah a state. Acceptance of this declaration by citizens of United States (i.e., collective intentionality) transformed this declaration into an institutional fact.

5 Shanks and Weber (2012) disagreed with Allen and March's (2012) interpretation. They refer to Bunge's (1977, p. 58) distinction between conceptual objects and substantial objects (or things) and posit that "the real world of substantial objects ultimately is unknowable. As a result, humans use conceptual objects to express their understanding of the real world" (Shanks & Weber, 2012 p. 968). In fact, "the only way humans can engage in discourse about or think about concrete things (and events that occur to concrete things) is via the concepts (conceptual models) that humans have devised to describe the things (and the events that occur to them)" (Shanks & Weber 2012, p. 968). Based on this argument, they believe that Bunge's ontology can indeed be used as a theoretical foundation for modeling conceptual objects using the example of the state of Utah (from March & Allen (2014)), based on Shanks and Weber's explanation, Bunge's ontology is indeed capable of representing a conceptual object such as the state of Utah (its recognition can be captured using the notion of an "event" in Bunge's ontology).

6 According to Soffer and Hadar (2007), different analysts create different models given the same domain. They defined model variations as "the differences in constructs and relations between adequately constructed models" (p. 599).

7 The keywords that we searched for were: "Ontology", "Bunge", and "Empirical".

8 Personal communication. Appendix B of (Burton-Jones et al. 2017) describes their data collection method. Their data from was cross-referenced with ours to make sure all the relevant papers are included.

9 We acknowledge that in practice, creation and interpretation of models are not mutually exclusive. Often people who create the model are also the ones who use them to improve their understanding of the domain as well as in the development of information systems. In our meta-analysis, however, subjects had no involvement in creation of the experimental material.

10 We incorporated the dependent variables separately from each other in our random-effects model. In other words, since we did not have access to the raw data from the studies in the pool, we could not determine the covariance of individual subjects' performance across different measures. Thus, using the random effects model, we treated the measures as independent from each other. Van den Noortgate, López-López, Marín-Martínez, and Sánchez-Meca (2015) have demonstrated that using the random effects model for analyzing multiple outcomes could lead to "appropriate mean effect size estimates, standard error estimates, and confidence interval coverage proportions in a variety of realistic situations" (p. 1274).

11 We do not claim memory recall is a poor measure by any means. Many of our academic examinations (with the purpose of measuring subject comprehension) have some free-recall components. As stated, we assume that practitioners will always have access to the conceptual models during the analysis and design phases (Parsons & Cole 2005).

12 Cohen's $d = (u_1 - u_2)/S_{\text{within}}$, where u_1 and u_2 are the sample means in two groups, and S_{within} is the within-groups standard deviation (Borenstein et al., 2011).

13 One of the papers included in our meta-analysis was the study done by Shanks et al. (2008), which was the topic of a discussion by Allen and March (2012) for not operationalizing the ontological guidance correctly. Later, Shanks and Weber (2012) provided a response to justify the validity of their experiment. We eliminated the study by Shanks et al. (2008) from our meta-analysis pool and reanalyzed the data. The tests with non-significant results remained non-significant, although with minor changes in the average effect size and confidence intervals. Our significant findings regarding the impact of ontological guidance on deep-level understanding remained unchanged as well, with average effect size of Cohen's d = 0.62, and 95% confidence interval of 0.18 to 1.07. Considering that Shanks et al.'s (2008) findings regarding deep-level understanding had the effect size of Cohen's d = 2.65, its removal reduced the variance and led to a narrower confidence interval.

14 In order to synthesize the study by Burton-Jones and Meso (2006), we only contrasted the performance of users in good and bad decomposition conditions (while ignoring the mediocre decommission condition).

15 As mentioned earlier, this hypothesis remained significant even after removing the study by Shanks et al. (2008), with average effect size of Cohen's d = 0.62, and 95% confidence interval of 0.18 to 1.07.

16 However, surface-level understanding of the domain reflects meaning as well as the context of usage, specifically, the application domain. In this case, the impact of ontological guidance on users' performance shows a rather large effect size of 0.91 (due to added expressiveness to the models), although, it was not statistically significant.

Arash Saghafi is an Assistant Professor at the Department of Management, Tilburg School of Economics and Management, Tilburg University, The Netherlands. His degrees include a BSc in Software Engineering from Sharif University of Technology, and MM, MScB, and PhD from University of British Columbia. His research has focused on data modelling and analytics, application of ontology in conceptual modelling, and empirical evaluation of design artifacts.

Yair Wand is Professor Emeritus in MIS at the Sauder School of Business, University of British Columbia, Canada. He received his D.Sc. in Operations Research from The Technion—Israel Institute of Technology, his M.Sc. in Physics from the Weizmann Institute—Israel, and B.Sc. in Physics from the Hebrew University, Jerusalem. His current research interests include theoretical foundations for information systems analysis and design, development and evaluation of systems analysis methods, and conceptual modelling.