

## EXPLORING THE EFFECTS OF EXTENSIONAL VERSUS INTENSIONAL REPRESENTATIONS ON DOMAIN UNDERSTANDING<sup>1</sup>

**Binny M. Samuel**

Carl H. Lindner College of Business, University of Cincinnati, PO Box 210130,  
Cincinnati, OH 45221-0130 U.S.A. {samuelby@uc.edu}

**Vijay Khatri and V. Ramesh**

Kelley School of Business, Indiana University, 1309 E. 10<sup>th</sup> Street,  
Bloomington, IN 47405 U.S.A. {vkhatri@indiana.edu} {venkat@indiana.edu}

*Cognitive research suggests that understanding the semantics, or the meaning, of representations involves both ascension from concrete concepts denoting specific observations (that is, extension) to abstract concepts denoting a number of observations (that is, intension), and vice versa. Consonantly, extant conceptual schemas can encode the semantics of a domain **intensionally** (e.g., ER diagram, UML class diagram) or **extensionally** (e.g., set diagram, UML object diagram). However, prior IS research has exclusively focused on intensional representations and the role they play in aiding domain understanding. In this research, we compare the interpretational fidelity of two types of representational encoding of cardinality constraints, an intensional schema using an ER diagram and its extensional analog using a set diagram. We employ cognitive science research to conceptualize that extensional representations will enable enhanced understanding as compared with intensional representations. Further, given that prior research suggests that the semantics of cardinality constraints remain challenging to understand, we focus on mandatory and optional cardinality constraints associated with relationships in these representations. Based on our laboratory experiments, we find that understanding with an extensional representation was (1) at least as good as that with an intensional representation for mandatory cardinality constraints and (2) significantly better for optional cardinality constraints. We also conducted an applicability check of our results via focus groups and found support for the perceived significance of extensional representations in practice. Overall, this research suggests that the tradition in IS research of exclusively focusing on intensional encoding of domain semantics should be reexamined.*

**Keywords:** Intensional representation, extensional representation, domain understanding, conceptual modeling, cardinality constraints, laboratory experiment, applicability check

<sup>1</sup>Andrew Burton-Jones was the accepting senior editor for this paper. Saonee Sarker served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly*'s website (<http://www.misq.org>).

## Introduction

How do individuals understand the domain semantics ascribed to representations? From the perspective of representations themselves, domain semantics can be encoded intensionally or extensionally (Mark and Roussopoulos 1987). For example, the intension of a database refers to the schema—the data structure—along with the regularities in the database (that is, the integrity constraints) that must be satisfied for every instance of a database value (Motro 1994). On the other hand, the extension of a database is a set of database values that populate the data structures. From the perspective of human cognition, research dating back to Aristotle (Frisch 1969) suggests that human understanding involves ascension from concrete concepts (that is, extension) to abstract concepts that describe a number of observations (that is, intension) (Roth and Hwang 2006); however, another point of view suggests that understanding involves moving from intension to extension (Ohlsson and Lehtinen 1997). Recent research in cognitive science recognizes the significance of both mechanisms (Whitney et al. 2009).

Given the importance of both intension and extension, it is not surprising that conceptual models have been created to help information systems (IS) professionals to produce intensional representations, such as UML class diagrams (Pilone and Pitman 2005) and ER diagrams (Chen 1976), as well as extensional representations, such as UML object diagrams (Pilone and Pitman 2005) and set diagrams (Elmasri and Navathe 2010).<sup>2</sup> Both types of representations (intensional and extensional) are regularly used in systems analysis and data management textbooks (Dennis et al. 2012; Elmasri and Navathe 2010; Gillenson 2012; Hoffer et al. 2015; Silberschatz et al. 2011; Umanath and Scamell 2015). However, prior IS research has overlooked the role extensional representations can play in domain understanding. Researchers have instead focused solely on intensional encoding of domain semantics using, for example, ontology diagrams (Bera et al. 2011), UML class diagrams (Allen and March 2012; Shanks et al. 2008), BPMN diagrams (Recker et al. 2011), and ER/EER diagrams (Bera et al. 2014; Bowen et al. 2009; Khatri et al. 2006). Our research takes a first step to examine the impact of semantics encoded using different types of representations on domain understanding by examining the following research question: What is the effect of the type of representation that is used to encode semantics—an intensional representation and its extensional analog—on domain understanding?

<sup>2</sup>In this paper, we refer to conceptual model as the grammar, and to conceptual schema as the representation created using a grammar. An example of a conceptual model and a conceptual schema is the ER model and an ER diagram, respectively.

To address our research question, we conducted experiments to compare domain understanding with a prevalent intensional representation and its extensional analog. In our experiments, we employed an ER diagram (Chen 1976) as a prevalent intensional representation and a set diagram (Elmasri and Navathe 2010) as its extensional analog. Given the significance of the semantics of the relationship construct in conceptual modeling (see Wand et al. 1999), we focus on the semantics of cardinalities: while the semantics of the cardinalities are encoded intensionally, for example, using the crow's feet notations in the ER diagram, they need to be satisfied extensionally, that is, for every instance of a database value. To clarify the semantics denoted by the crow's feet notations, these constraints are sometimes also shown extensionally in database textbooks using a set diagram. Within cardinalities, we concentrate on the semantics of minimum cardinalities only, which are characterized as optional or mandatory (Chen 1976; Thalheim 1992). Optional and mandatory cardinalities are known to present different levels of cognitive difficulty to understand (see, for example, Bodart et al. 2001; Gemino and Wand 2005; Shanks et al. 2008). For optional and mandatory cardinalities represented in an ER and a set diagram, we employed several dependent measures to gauge understanding of the domain semantics. We found that understanding with a set diagram was (1) at least as good as with an ER diagram for mandatory cardinality constraints and (2) significantly better for optional cardinality constraints.

The rest of this paper is structured as follows. We first introduce background material relevant to this research and then present our proposition. After providing an overview of our experiments, we describe each experiment in detail. Next, we present a discussion of our results. We round out this paper with an applicability check (from practice) and the implications of our findings to research and practice.

## Background

The basis for intensional and extensional representations is rooted in cognitive science. In this section, after presenting that foundation for this research, we present an example of an intensional and extensional representation, respectively, in conceptual modeling. Finally, we frame our research, which focuses on relationship cardinalities, in the context of prior research that has examined the semantics of relationships.

### *Intensional Versus Extensional Representations*

Dating back to Aristotle, there has long been a tradition to differentiate between semantics that are ascribed to represen-

tations intensionally and extensionally (Zhou and Mao 2010). According to the classical and prototype theories of concepts, the meaning of a concept is characterized by its gist or properties (i.e., intension). The classical theory considers a set of necessary and sufficient properties in ascribing concept meaning (Rosch 1973). Similarly, the prototype theory (Rosch 1973) suggests that people tend to focus on central tendencies of a concept (e.g., its essential properties) rather than its individual members (Posner and Keele 1970). On the other hand, exemplar theory suggests that the meaning of a concept is determined by a set of instances of things that exist in the referring domain (Nosofsky 1991), or its extension.

Prior cognitive science research has debated which representation of concepts is better for understanding. Some research suggests that individuals demonstrate better understanding with extensional than intensional representations (Koedinger and Nathan 2004; Markovits and Vachon 1990). Individuals exist in a world populated with concrete things that are a part of their everyday life. Extensional representations provide an air of familiarity that connect to the way individuals function in everyday life via encounters and interactions with concrete things in the world (Bunge 1977; Koedinger et al. 2008); extensional representations thus make concepts more coherent to individuals (Anderson et al. 1976). Similarly, extensional representations also rely less on specific terminology and symbols that have to be translated to be understood (e.g., symbols such as the legends on a map or the crow's feet notation in an ER diagram), and instead employ depictions that are closer to real world things, thus eliminating some of the translation and interpretation an individual might have to make for understanding the real world (Koedinger and Nathan 2004). Finally, extensional representations often contain ready-made examples that individuals can utilize for understanding instead of having to create their own (Atkinson et al. 2000).

Other research suggests the opposite. Intensional representations often utilize specialized terminology or symbols to simplify the complexity of the world it represents to its essence (Koedinger and Nathan 2004).<sup>3</sup> By reducing the complexity of the world into its common essential properties, intensional representations reduce the cognitive burden on an individual's working memory (Koedinger et al. 2008). Specifically, individuals can focus on fewer higher-level concepts described by their common essential properties instead of all the things that are referents of the intensional representation. This reduction of cognitive burden has been referred to as cognitive economy (Lakoff 1987; Rosch 1978). Due to the focused amount of information presented in them, intensional

representations lead to fewer distractions from potentially irrelevant information for understanding (Moreno et al. 2011).

In summary, prior cognitive science research suggests that both extensional and intensional representations can play a role in understanding the semantics of the real world. While the former, extensional representations, employ less terminology, the latter, intensional representations, can alleviate the cognitive burden associated with the complexity of the world.

### ***Intensional Versus Extensional Representations in Conceptual Modeling***

The starting point for this research was the extant use of intensional and extensional representations in the practice of systems analysis and design (Kuske et al. 2002). In this research, an *intensional representation*, which is characterized by a set of properties of a concept, embodies abstract encoding of concepts in the real world; the ISO (International Standards Organization) terminology standard 1087 (OMG 2015) defines it as "the set of characteristics which make up the concept." On the other hand, an *extensional representation*, which is frequently represented by a set of concrete things, establishes the referential encoding of a concept in the real world; ISO 1087 defines it as "the totality of objects to which a concept corresponds." As noted earlier, both types of representation are employed in systems analysis and design; see, for example, UML class diagrams and ER diagrams vis-à-vis UML object diagrams and set diagrams.

The role of extensional representations has been recognized in the design of programming languages, agile development, and conceptual modeling. For example, in query programming languages for object-oriented databases, prior research has unified the intensional (e.g., a query to retrieve relationships between classes) and extensional (e.g., a query to retrieve individual objects) aspects of the language (see, for example, Kifer and Lausen 1989; Lecluse et al. 1988; Savnik et al. 1999). In the case of agile development, the use of personas, detailed narratives about plausible instances of software users, is said to not only improve communication about target users but also increase the focus on them (Ambler 2014a; Cooper 2004; Ma and LeRouge 2007). In the context of this research on conceptual modeling, extensional representations can play one of two roles:

- (1) An extensional representation can help establish the referential meaning of the real world by presenting a representation that maps closely with how the world filled with concrete things is perceived. For example, a UML object diagram is employed in the practice of systems analysis and design as an *extensional analog* of a UML class diagram (Ambler 2014b) as it employs con-

<sup>3</sup>Koedinger and Nathan used grounded math word problems versus abstract math equations. Their conceptualization of abstract and grounded representations is consonant with the notion of intensional and extensional representations, respectively.

crete things to establish the referential meaning of the real world by showing one instance of each attribute represented in the classes.

- (2) An extensional representation can explicate the semantics of a real-world domain. For example, a set diagram, an *extensional analog* of an ER diagram, can explicate the allowed regularities, or constraints, that must be enforced for every instance of a database value. Our research focuses on the second role that extensional representations can play in conceptual modeling.

Prior conceptual modeling research has focused on schema creation or schema understanding. During *schema creation*, conceptual modeling research has recognized the role of both intension and extension. For example, prior research suggests that individuals who make frequent and orderly cognitive transitions between an intension and an extension of a domain are able to create more accurate conceptual schemas (Srinivasan and Te'eni 1995). Additionally, the discovery of meaningful concepts (e.g., classes, entity types) that are used to depict an intensional encoding of a conceptual schema are best derived from the extensions known by domain experts (Parsons 1996; Parsons and Wand 1997, 2008).

However, conceptual modeling research on *schema understanding* (the focus of our research) has exclusively focused on intensional representations. We reviewed empirical research on schema understanding in top-tier IS publication outlets (i.e., *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, and *Journal of the Association for Information Systems*) with respect to the type of representation (intensional versus extensional) employed to encode the semantics associated with the substance and form of a domain. We found that prior research on schema understanding has exclusively focused on intensional representations (see Bera et al. 2014; Bodart et al. 2001; Burton-Jones and Meso 2006, 2008; Dunn et al. 2011; Khatri et al. 2006; Parsons 2002, 2011; Shanks et al. 2008). Thus, little is known about the role extensional representations can play in domain understanding. Given that prior research in schema creation recognizes the role of the extension of a domain and that Chen (1976), in a seminal paper on the most popular conceptual model, acknowledged the role of extensional representations (e.g., entity set, relationship set, value set) in domain understanding, we believe that it is important to examine the potential role of extensional representations in domain understanding.

## Understanding Relationships

This research focuses on the semantics of cardinalities that are associated with an interaction relationship, and the role of

intensional and extensional representations in facilitating the understanding of these semantics (i.e., domain understanding). Three prior studies in conceptual modeling have theoretically and empirically examined the semantics that are ascribed to different types of relationships such as interaction and part-whole. First, based on the examination of the interaction relationships, Bodart et al. (2001) employed semantic network theory (Collins and Quillian 1969), which suggests human semantic memory can be thought of as a network of nodes (i.e., concepts) linked by pathways. They found that optional cardinalities were not strong stimulations for activating pathways and therefore limit deeper processing (i.e., understanding) of conceptual schemas. Second, Gemino and Wand (2005) examined interaction relationships in conceptual schemas using the cognitive theory of multimedia learning (CTML) (Mayer 2001). CTML suggests humans have two separate information processing channels (i.e., visual and verbal) available for understanding representations, and conceptual schemas with optional interaction relationships require more cognitive resources than schemas without them. Third, Shanks et al. (2008) examined how to best express the semantics of a part-whole, or a composite, relationship in conceptual modeling, which has sometimes been depicted via a relationship among entity types. The theory of ontology proposed by Bunge (1977) and the theory of ontological clarity (Wand and Weber 1993) were utilized to suggest that classifying phenomena in a real world domain in accordance to the structure of the world will aid in understanding the part-whole construct.

Each of the aforementioned studies used laboratory experiments with various dependent measures to establish their findings on schema understanding of relationships. Bodart et al. had individuals rely on their memory to answer true/false questions about the rules of a domain depicted in a (conceptual) schema. Individuals also had to rely on their memory to answer inference questions that required logical explanations about a domain event (e.g., reasons for why a project is delayed). Gemino and Wand employed fill-in-the-blank (about the rules of a domain depicted in a schema) and inference questions; both dependent measures used memory recall of the schema. Finally, Shanks et al. (2008) asked individuals to complete problem solving questions based on domain events. Individuals first indicated if the semantics of the schema indicated the event was possible or not (they could also indicate "not sure"), and then provided their rationale. Individuals had access to the schema when answering these questions.

In this research, we further this tradition of developing normative approaches in conceptual modeling using a variety of dependent measures. We focus on assessing the potential efficacy of extensional encoding of the interaction relationship cardinalities. In examining relationships in conceptual

modeling, prior research suggests that representing domain semantics associated with cardinalities is one of the most important aspects (Currim and Ram 2012; Liddle et al. 1993; Olivé 2007; Ram and Khatri 2005; Rosca et al. 2002; Rundensteiner 1992; Rundensteiner et al. 1994). Prior empirical research has examined different aspects related to the semantics of cardinalities. Some of this research has compared different forms of intensional representations for verifying cardinalities in a conceptual schema. For example, Juhn and Naumann (1985) found that individuals could better identify cardinality constraints in graphical as opposed to verbal (e.g., text) representations. Dunn et al. (2011) also examined different types of representations to uncover the one that was the easiest for individuals to find cardinality discrepancies in a conceptual schema. While the aforementioned cardinality-related research has provided sound prescription for practice, it does not delve into examining the role that extensional encoding of cardinalities can play in facilitating the understanding of relationship-related domain semantics.

In summary, prior IS research has exclusively focused on representations that intensionally encode domain semantics using a variety of dependent measures. Second, prior research suggests that the semantics of cardinalities are challenging to understand in conceptual schemas (Batra et al. 1990; Bodart et al. 2001; Dunn et al. 2011; Gemino and Wand 2005; Wand et al. 1999). Third, while aspects such as optionality have been examined in the context of the interaction relationship, cardinalities, which are integral to the semantics of the relationship construct, have not received adequate attention.

## Theory and Proposition

In this section, we address the aforementioned issues in prior research by developing a proposition regarding the benefit of extensional representation on individuals' understanding of cardinality-related semantics. To scope the work, we focus on the ER diagram as the focal intensional representation and the set diagram as the focal extensional representation. We focus on these representations because the ER diagram is a well-known and widely used intensional representation (Davies et al. 2006; Fettke 2009) and the set diagram, whose origin traces back to Chen (1976), is its prevalent extensional analog (Elmasri and Navathe 2010).

### Illustrations of Intensional and Extensional Conceptual Schemas

Figure 1 illustrates an example of intensional encoding of the semantics of a mandatory cardinality constraint (based on

Bodart et al. 2001). Meanwhile, Figure 2, which presents the extensional analog of Figure 1, explicates cardinality-related regularities or constraints that must be enforced. The figures show two entity types, RESEARCH PROJECT and PROJECT REPORT. In Figure 1, the RESEARCH PROJECT and PROJECT REPORT entity types have seven (e.g., Start Date) and three attributes (e.g., Title), respectively, which are represented by labeled ovals. The identifier attribute that uniquely identifies each entity instance, for example, Project ID in RESEARCH PROJECT, is represented by underlining the attribute label. The relationship type Describes represents the association between RESEARCH PROJECT and PROJECT REPORT. The associated cardinality constraint indicates that a research project (an entity instance) must be described by at least one, but could have many, project reports (shown using crow's feet notation as  $1 \leq \infty$ ). From a project report's point of view, the cardinality constraint indicates that a project report (entity instance) describes exactly one research project (shown using crow's feet notation as  $1-1$ ). Note that while the crow's feet notation is employed to encode domain semantics intensionally (in the schema in Figure 1), the semantics of this constraint apply to the extension of the conceptual schema (that is, to the instances).

Figure 2, which presents an extensional analog of Figure 1, explicates the cardinality-related semantics extensionally. The entity type, for example, RESEARCH PROJECT, along with the attributes (in parentheses after the entity type) are indicated below the rectangle. Within each rectangle are a few filled-in circles, each of which serves as a representation of an entity instance. Each entity instance is shown with concrete values for their corresponding attributes. For example, the RESEARCH PROJECT entity type has three entity instances with values for ProjectID as follows: 77, 79, and 16. Similarly, the PROJECT REPORT entity type has four entity instances with values for Report# as follows: 95, 103, 107, and 88. We note that instances in this current research are only shown as representative members of an intensional representation. Given the focus of this research on cardinalities for interaction relationships, our emphasis is on exemplars (Nosofsky 1991) for the cardinality constraints which we explain next.

The relationship type Describes is represented with a diamond in Figure 2. The relationship type contains filled-in circles that serve as a representation of the relationship instances, and each relationship instance has two associated entity instances. In Figure 2, the example relationship instances connect entity instances for RESEARCH PROJECT and PROJECT REPORT. In the extensional representation, the semantics of the cardinality constraints are inferred from the relationship instances associated with the entity instances. Specifically, one can infer the maximum and minimum cardinalities based on the maximum and the minimum number of relationship instances asso-

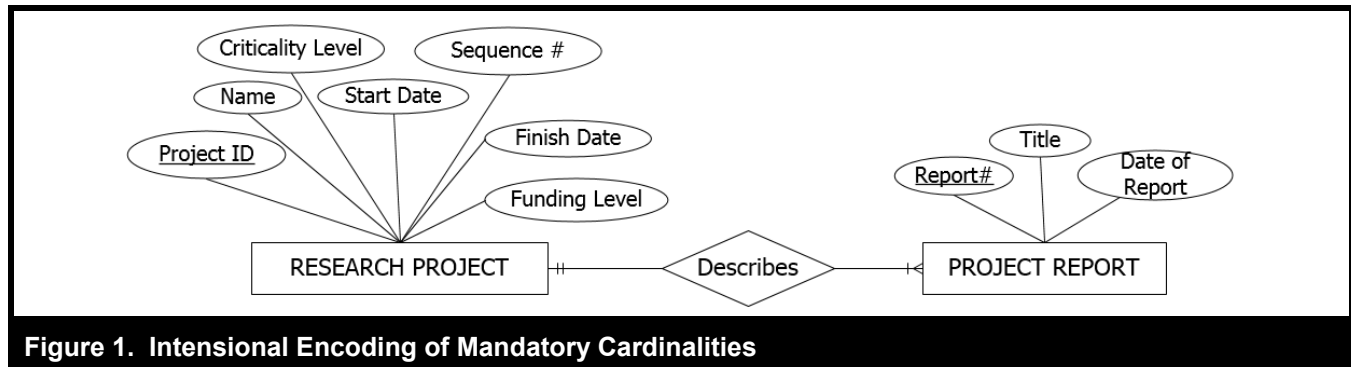


Figure 1. Intensional Encoding of Mandatory Cardinalities

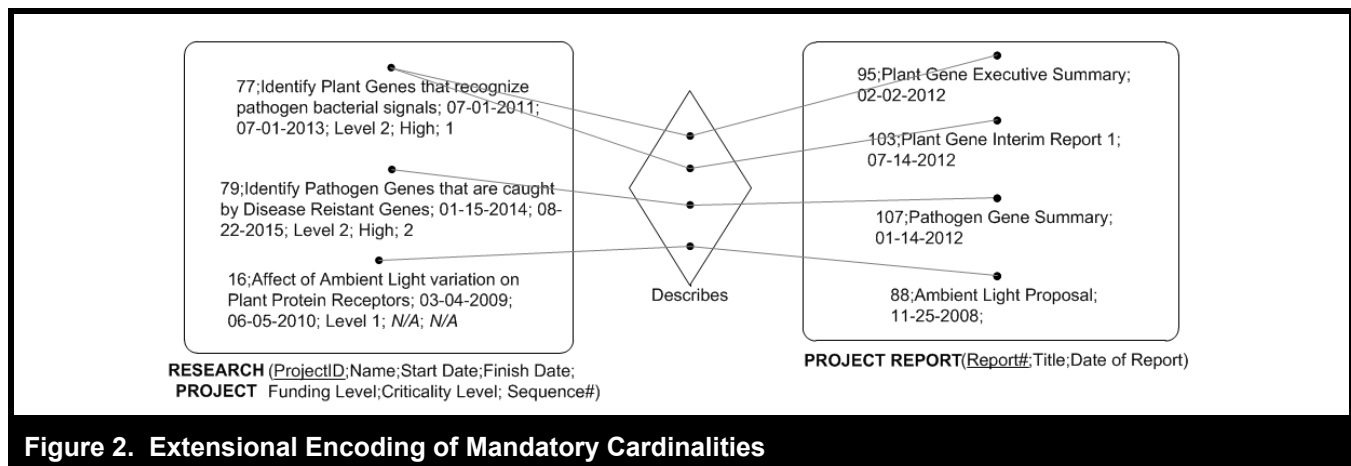


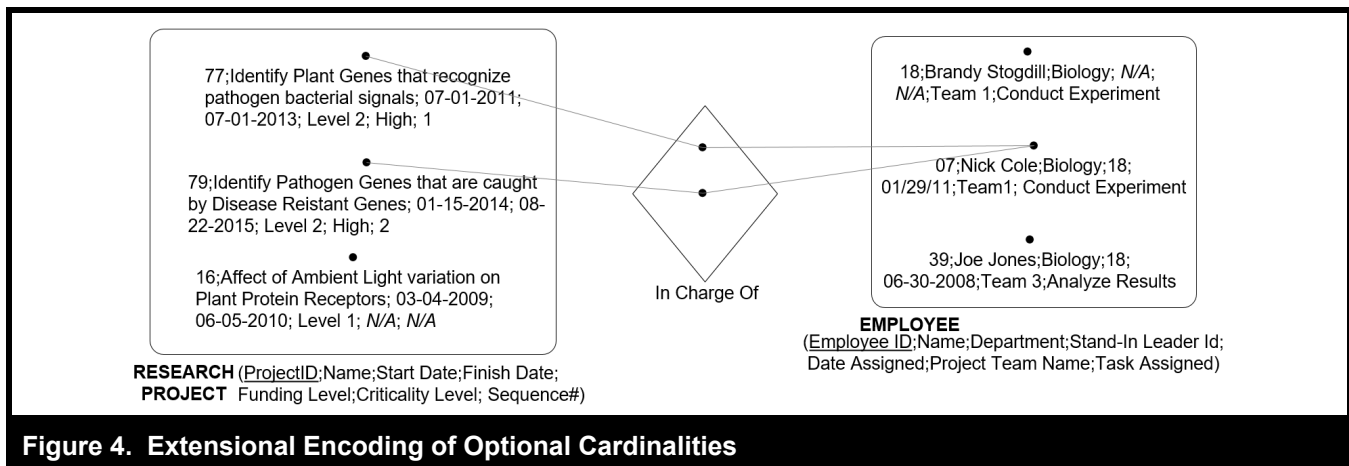
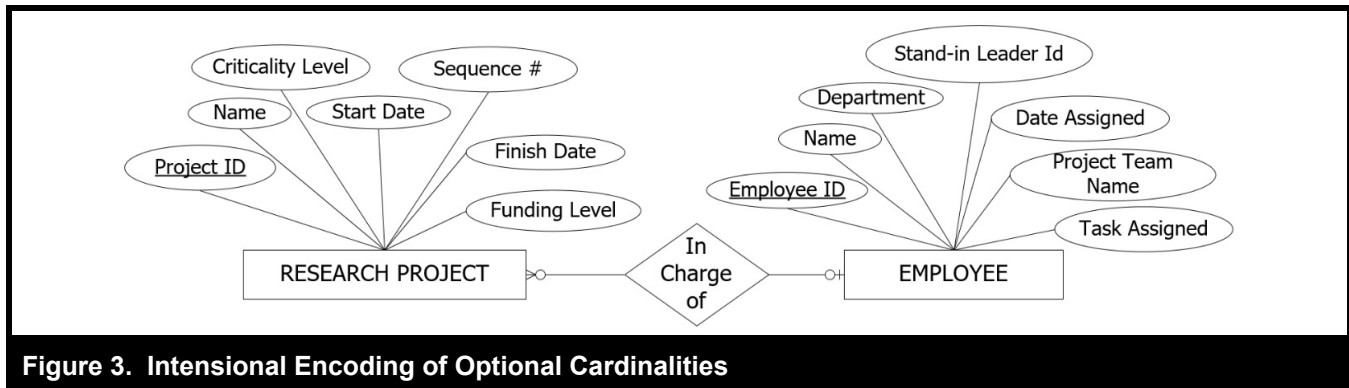
Figure 2. Extensional Encoding of Mandatory Cardinalities

ciated with representative entity instances. The examples in Figure 2 indicate that RESEARCH PROJECT 77 (entity instance) has multiple relationship instances in Describes, thus the associated maximum cardinality for the entity type is “many.” The relationship instances also indicate that every RESEARCH PROJECT is associated with a relationship instance in Describes which signifies the minimum cardinality of “one” (also referred to as the mandatory relationship) for the entity type. Figure 2 also shows that the PROJECT REPORT entity instances have both a maximum and minimum relationship instance of one in the Describes relationship, thus the maximum and minimum cardinality are both one. In ascribing the semantics of cardinalities, it is difficult to cope with the extension of concepts in their entirety and that is why our extensional representations employ exemplar typicality to encode the meaning of cardinalities. Such a representation allows inferences to be made about the meaning of the cardinality constraints among the entity instances for a specified relationship.<sup>4</sup>

<sup>4</sup>Our operationalization of encoding cardinality-related semantics assumes that exemplar entity and relationship instances sufficiently represent the semantics of cardinalities.

Figures 3 and 4 illustrate the intensional and extensional representations, respectively, of two optional cardinality constraints. In Figure 3, the use of optionality for the minimum cardinality indicates that an instance of EMPLOYEE does not have to be In Charge of an instance of RESEARCH PROJECT ( $\geq 0$ ). Similarly an instance of RESEARCH PROJECT does not need to have an Employee In Charge Of ( $0 +$ ) it.

In Figure 4, the RESEARCH PROJECT instances are associated with at most one instance in the In Charge Of relationship, thus, the maximum cardinality is “one.” It also indicates that *not* every RESEARCH PROJECT instance has an associated instance in the In Charge Of relationship, thus, signifying the minimum cardinality of “zero.” Figure 4 also shows the EMPLOYEE instances can have a maximum of many instances in the In Charge Of relationship (see EMPLOYEE 07); therefore, the maximum cardinality is “many,” and not every EMPLOYEE has an associated instance in the In Charge Of relationship (see EMPLOYEE 18 and 39), thus, the minimum cardinality is “zero.”



### Proposition Development

In comparing an intensional representation with its extensional analog, we first highlight two aspects that intrinsically characterize the representations in this research. First, for the extensional schema, the cardinality semantics can be inferred by observation of the relationship instances whereas the intensional schema denotes cardinality constraints solely using a specific syntax (Zhou and Mao 2010). Both types of representations require inferring domain semantics: via examination of the relationship instances in the former and via observation of the cardinality syntax in the latter. Second, more elements are depicted in our extensional schema than our intensional schema, which is consistent with prior cognitive science research which suggests that an intensional representation contains less perceptual information than an extensional one (Kaminski and Sloutsky 2012). Thus, while extensional schemas signify the meaning without the use of cardinality symbols, an individual needs to process more information. From a cognitive perspective, we assess the effect of the tradeoffs between the representations (intensional and extensional) on domain understanding.

Based on cognitive science research, we postulate that individuals will demonstrate better understanding of the cardinality constraints when the constraints are encoded extensionally as opposed to intensionally. First, while the cardinality constraints are expressed succinctly in an intensional conceptual schema (e.g., using the crow's feet notation), the extensional representation encodes the real-world semantics more directly (i.e., via relationship instances). Prior research suggests that providing examples (e.g., our set diagram instances) helps with understanding representations (Beveridge and Parkins 1987; Day and Goldstone 2012; Glenberg et al. 2004; Goldstone and Medin 1994). Although the semantics of the cardinalities are explicitly depicted via cardinality symbols in an intensional representation, individuals may not understand the semantics represented by the symbols (Koedinger and Nathan 2004; Lave 1988). Even in cases where an individual does seem to know the encoded meaning of the cardinality syntax, individuals perform better with examples such as those found in an extensional representation (Allen and Brooks 1991). Therefore we posit that extensional representations make the domain semantics of cardinality constraints more concrete (see Paivio 2007) and thus more coherent (Anderson et al. 1976).

Second, the properties of a concept need to be instantiated to be understood (Hahn and Chater 1998). In particular, cardinality constraints apply to the extension of the conceptual schema, thus implying that understanding the semantics of cardinalities in an intensional schema requires an individual to interpret the cardinality notation (e.g., crow's feet notation) to their extension. Prior research suggests that when a representation's information is encoded in a manner that is consistent with the task, it will match the kind of cognitive processes the individual will use for task completion, and thus the task solution will be facilitated (Adams et al. 1988; Lockhart et al. 1988). Given that cardinality symbols in an intensional schema have to be translated to be understood, we believe that an extensional schema will result in better performance as it intrinsically matches the encoding necessary for understanding.

Third, intensional representations lack ready-made examples that individuals can utilize for understanding, and instead individuals will have to create their own. Providing ready-made examples has been shown to be beneficial for individuals to better understand information and they are less likely to commit conceptual errors when tested on their understanding (Atkinson et al. 2000; McNeil et al. 2009). These ready-made examples can provide useful checks and constraints that prevent certain kinds of mistakes (Koedinger et al. 2008). Thus individuals will demonstrate better understanding with an extensional conceptual schema that contains ready-made examples than with an intensional conceptual schema that lacks these examples.

Based on the rationale highlighted above, we offer the following proposition:

*Proposition:* Individuals given extensional encoding of cardinality constraints in a conceptual schema will demonstrate a more accurate understanding of the domain semantics than those given intensional encoding of cardinality constraints.

Given that prior research that has examined the meaning of relationships has found mandatory and optional relationships to present different levels of cognitive difficulty (Bodart et al. 2001), we test our proposition for both types of cardinality constraints.

## Overview of Experiments

We conducted laboratory experiments to assess the proposition. In the following, we present details of the materials that were employed. We then describe the pre-test that was conducted. Finally, we characterize the participants and

outline the general procedure employed in the experiments. Guidance in the development of the methods to assess the proposition is based on the work of Burton-Jones et al. (2009), who provide recommendations for empirical evaluation of conceptual modeling research. We compared the interpretational fidelity of intensionally encoding the semantics of cardinality constraints, via the ER diagram, and its extensional analog, the set diagram. Interpretational fidelity refers to the accuracy of the alignment between an individual's understanding of the semantics of the cardinalities in a representation compared to the correct interpretation of those semantics.

## Materials

We developed three sets of materials for this experiment. Please refer to Appendices A–E for the materials discussed in this section.

### Training Material

The first set of materials were training videos that showed the participants how to employ a conceptual schema to understand domain semantics. Participants only received training for the type of representation—intensional or extensional—they would use during the experiment. Participants in an intensional treatment were trained to understand the components of an ER diagram while those in an extensional treatment were trained to understand the components of a set diagram. The training videos were both approximately 10 minutes in length. To ensure that our participants understood the material in the training video, they were asked to complete six training questions. All participants completed the training and correctly answered all training questions which suggests that they understood the material presented.

### Conceptual Schemas

The second set of materials was one of three conceptual schemas that participants used during the experiments. We used the ER model to develop a schema (ER diagram) to intensionally encode the cardinality semantics with the crow's feet notation, the most commonly used cardinality notation in industry (Moody 2009). The set notation was used to develop a schema (set diagram) that extensionally encodes cardinality semantics. The aforementioned schemas allowed us to assess our proposition.

We also developed an alternative extensional representation with a larger number of instances (compared to the set



diagram), referred to as the set+ diagram, to assess differences in understanding between two extensional representations.<sup>5</sup> We used an increased number of instances for each entity type in this alternative representation. This allowed us to increase the number of optional instances for optional cardinality, thus showing a stronger pattern of optionality. We could therefore assess the efficacy of stronger optionality patterns on our dependent measures. Similarly, to encode mandatory cardinality, all entity instances participated in a relationship, thus strengthening the emphasis of the mandatory cardinality constraint.

The same sales domain was chosen for all schemas and they included constructs to represent MANUFACTURER, PRODUCT, PRODUCT LINE, CUSTOMER, ORDER, etc.; see Appendices A–C. Our emphasis was on comparing the effects of intensional versus extensional encoding of cardinality semantics, therefore unlike prior research (Bodart et al. 2001; Gemino and Wand 2005) each of our schemas had both optional and mandatory relationships.

## Dependent Measures

Prior research in cognitive science suggests that the benefit of any representation is tied to the way in which it is used (Hahn and Chater 1998), hence we employed several dependent measures. As shown in the first two columns of Table 1, prior research in conceptual schema understanding has employed different types of dependent measures to assess understanding of the domain from a schema. Saghaei and Wand (2014) characterize dependent measures as surface- or deep-level. Dependent measures that assess *surface-level understanding* employ working memory only and focus on the factual knowledge about the constructs of a conceptual model. In prior literature, surface-level understanding has been examined using tasks that require *syntactic* and *semantic comprehension* as well as those that entail *elaborative reconstruction*. In contrast to surface-level understanding, *deep-level understanding* seeks to gauge the ability of a problem solver to use the information in a conceptual schema to reason about the domain; such tasks require integrating information from the schema with that in long-term memory. Deep-level understanding has been assessed using tasks that employ *problem solving*. The delineation of surface-level understanding versus deep-level understanding is common in the IS schema understanding literature and also has its roots in cognitive science (see Alexander 1997).

<sup>5</sup>We originally designed the study as a two-group study (i.e., set diagram versus ER diagram). During the review process, an alternative representation of the extensional representation (set+ diagram), which employed stronger cardinality patterns, was suggested. Therefore, we included another group in our design.

Our experiment 1 compared domain understanding using the ER, set, and set+ diagrams, respectively. We employed three measures of understanding (see Table 1) for both optional and mandatory cardinality constraints. Experiment 2 compared understanding of an ER diagram with a set diagram for optional and mandatory cardinalities using two different dependent measures than those used in Experiment 1. Overall, we employed five dependent measures across the experiments to gauge understanding of

- (1) *semantic comprehension* using the following tasks:
  - (a) fill-in-the-blank, (b) one-relationship report, and (c) two-relationship report;
- (2) *elaborative reconstruction* using the memory recall task; and
- (3) problem solving using the *schema-based problem-solving task*.

Our training questions were syntactic comprehension questions and were considered a baseline for understanding the domain semantics of the schemas. Some of our dependent measures were adapted from prior research; however, the report questions, which required a respondent to navigate one or two relationships in the schema, are novel. Given that a report, or a query, is merely an operational way of making domain semantics explicit, these report questions gauged the ability of an individual to unambiguously understand the semantics of the domain which may be implicit in the schema.

## Pre-Test of Materials

A pre-test with six individuals among the treatment groups was conducted to validate the experimental materials as well as to look for situations during the experiment where a participant might have difficulty in understanding the experimental materials. This pre-test also sought to gain feedback on the effectiveness of the training employed at the start of the experiment. The qualitative nature of the pre-test allowed us to better understand the potential sources of errors due to misinterpretation of the experimental material during conceptual schema understanding across different experimental treatments (Tashakkori and Teddlie 1998). Participants in the pre-test, who were drawn from the same pool as the participants of the experiment, were trained and instructed to think aloud while answering the questions. In addition to providing an answer, our pre-test participants had to explain how they arrived at their answer. The pre-test thus helped increase internal validity with respect to the experiment design and materials.

**Table 1. Dependent Measures Used in Prior Research on Schema Understanding and in This Research**

| Dependent Measure  | Example   | Experiment #: Representations Employed  | Operationalization of the Dependent Measure                             |
|--|---|---|---|
| <i>Syntactic</i> comprehension assesses whether an individual understands the vocabulary of the constructs used to create a conceptual schema (e.g., a diamond in the ER grammar refers to a relationship) (Bodart et al. 2001; Khatri et al. 2006; Kim and March 1995). | What is the minimum:maximum cardinality of the relationship between PRODUCT and ORDER?<br>(a) 0:M and 0:M; (b) 1:M and 0:M; (c) 0:M and 1:M; (d) 1:1 and 0:M (Kim and March 1995)   | <i>Experiments 1 and 2:</i> Set Diagram vs. ER Diagram vs. Set+ Diagram in Exp. 1; Set Diagram vs. ER Diagram in Exp. 2 | Syntactic (training questions)  |
| <i>Semantic</i> comprehension assesses whether an individual understands the meaning or semantics of the constructs of a conceptual model in a conceptual schema (Bodart et al. 2001; Khatri et al. 2006).   | A SALES PERSON is responsible for (a) exactly one PRODUCT LINE; (b) at the most one PRODUCT LINE; (c) no more than one PRODUCT LINE; (d) zero or more PRODUCT LINES (Khatri et al. 2006)<br><br>A warehouse has in it a minimum of _____ product lines. | <i>Experiment 1:</i> Set Diagram vs. ER Diagram vs. Set + Diagram   | Fill-In-the-Blank<br>One Relationship Report<br>Two Relationship Report |
| <i>Elaborative reconstruction</i> that examines processing of the domain semantics has employed memory recall (Bodart et al. 2001).  | Recall of entities, relationships, etc. (Bodart et al. 2001)  | <i>Experiment 2:</i> Set Diagram vs. ER Diagram   | Memory Recall   |
| <i>Schema-based problem-solving</i> requires an individual to more deeply apply the information represented in the schema compared to semantic comprehension (Khatri et al. 2006; Shanks et al. 2003).   | A team leader has resigned. Does the model allow the team to continue to work on the project without him? (Shanks et al. 2008)  | <i>Experiment 2:</i> Set Diagram vs. ER Diagram   | Schema-Based Problem-Solving  |

## Participants

The population employed was part of an established subject pool at a U.S.-based university business school. All of the participants were enrolled in a required introductory business analytics course for all business students. The course, which is typically taken by students in their second year of undergraduate education, focuses on introductory business analytics using Microsoft Excel.<sup>6</sup> Students participating in the study were awarded extra credit equivalent to one percent of their total grade in the course. The extra credit was intended to motivate students to participate, but not reward them for their performance.

<sup>6</sup>The course has a prerequisite course that provides an introduction to (1) using databases (i.e., Microsoft Access) with emphasis on data retrieval via query by design and (2) basic data analysis using Microsoft Excel.

We developed a questionnaire to gather background information about our participants (see Appendix D). The questionnaire asked each participant to provide a subjective skill rating on measures used in prior conceptual schema understanding research (e.g., the ability to use conceptual schemas). Our participants' responded to the questions via a Likert scale with a numeric value between 1 (i.e., strongly disagree) and 7 (i.e., strongly agree). We note that, across all of our experiments, our participants reported an average of 4.02 with respect to their skill of using conceptual schemas with no significant differences between the groups. While our participants were not practitioners with expertise in creating and using conceptual schemas, their self-reported skill is similar to self-reported measures found conceptual modeling research (Bera et al. 2011, 2014; Burton-Jones and Meso 2006).

## General Procedures

Participants arrived at a computer laboratory and started the session on a desktop computer with a web browser that connected to a website with the training material and the experimental tasks. The participants were randomly assigned to an experimental treatment group upon arrival at the laboratory as this assignment allowed them to watch a corresponding training video for their treatment. Participants then completed a background questionnaire that enabled us to gather demographic and background information (see the previous subsection). Next a lab administrator distributed a printed conceptual schema for each participant's treatment group. The lab administrator had to enter a password on the website for the participants to advance to the experimental task and this ensured each participant received the appropriate treatment diagram. The participants then proceeded to use a computer to complete the experimental tasks.

## Experiment 1

We first present the design and then elucidate the dependent measures employed in experiment 1. We next describe our findings for experiment 1.

### Experiment 1 Design

Experiment 1 assessed the differences in interpretation fidelity of an extensional (set diagram) and an intensional (ER diagram) representation, and the differences in interpretation fidelity introduced by an alternative extensional representation referred to as the set+ diagram. There was one between-subjects factor, the type of representational encoding of the semantics in a conceptual schema. Participants received only one of the schemas described earlier. This allowed us to examine the effects of the type of representation on the participant's ability to understand the domain semantics. The dependent measure was *understanding* of the semantics of mandatory and optional cardinality constraints.

A total of 147 participants (49 in each treatment group) were recruited for this research from a population of undergraduate business students and randomly assigned to one of the three between-subjects treatment groups: set diagram, ER diagram, and set+ diagram. To examine if we had successfully randomized our participants across the groups with respect to their backgrounds, we conducted *F*-tests. Our findings suggest no significant differences between skill levels across the treatment groups. Also, a Pearson chi-square test ( $\chi^2 = 0.19$ ,  $df = 2$ ) and likelihood ratio test ( $\chi^2 = 0.19$ ,  $df = 2$ ) suggested

that there was no association between the treatment group assignment and the participant's gender.

### Experiment 1 Dependent Measures

We employed the same dependent measures to gauge domain understanding across the treatment groups: (1) fill-in-the-blank; (2) one relationship report; and (3) two relationship report (see Appendix E5). First, fill-in-the-blank consisted of six questions, of which three each were related to optional and mandatory cardinalities. Second, based on the number of relationships types that needed to be examined, there were two kinds of relationship report questions. There were six of each kind that required a respondent to evaluate (1) only one relationship between two entity types; (2) two relationships traversing three entity types. For each block of six questions, half were related to optional and the other half to mandatory cardinality constraints. The order of questions was randomized to include both optional and mandatory cardinality constraints. We ensured that both representations could be used to answer each question.

### Experiment 1 Analysis and Findings

We examined understanding of cardinality-related semantics by participants using their performance (percent correct) on our three dependent measures: (1) fill-in-the-blank; (2) one relationship report; and (3) two relationship report. We report our results in Table 2 where we first show the mean accuracy for performance on all dependent measures by treatment group (i.e., set diagram, an ER diagram, or the set+ diagram), and the findings of our planned comparisons between our diagram treatment groups.

Comparisons between the set versus ER diagrams indicates our participants were more accurate with the set diagram than with the ER diagram for dependent measures that assessed the understanding of the semantics of optional cardinalities. On the other hand, for mandatory cardinalities, the difference in accuracy between the two groups was not statistically significant.

All differences between set+ versus ER diagrams and set+ versus set diagrams were not statistically significant. Much like the set diagram, the set+ diagram edged past the ER diagram for optional cardinality understanding. The ER diagram, however, was in general better than the set+ diagram for mandatory cardinality understanding with one exception (the two relationship report). Finally, those with the set+ dia-

**Table 2. Detailed Results for Experiment 1**

| Task                     | Mean Accuracy (% correct)        |                                 |                                   | Mean Differences<br>Set vs. ER<br>(Std. Error) | Mean Differences<br>Set+ vs. ER<br>(Std. Error) | Mean Differences<br>Set vs. Set+<br>(Std. Error) |
|--------------------------|----------------------------------|---------------------------------|-----------------------------------|--|---|--|
|                          | Set Diagram<br>49<br>respondents | ER Diagram<br>49<br>respondents | Set+ Diagram<br>49<br>respondents |  |   |  |
| Fill-in-the-Blank        |                                  |                                 |                                   |  |   |  |
| Optional                 | 75.5                             | 59.2                            | 61.2                              | 16.3*<br>(7.5)                                 | 2.0<br>(8.0)                                    | 14.3<br>(7.8)                                    |
| Mandatory                | 83.7                             | 78.2                            | 76.9                              | 5.4<br>(6.1)                                   | -1.3<br>(6.3)                                   | 6.8<br>(6.3)                                     |
| One Relationship Report  |                                  |                                 |                                   |  |   |  |
| Optional                 | 50.3                             | 32.0                            | 46.3                              | 18.4*<br>(7.9)                                 | 14.3<br>(8.0)                                   | 4.1<br>(8.4)                                     |
| Mandatory                | 68.7                             | 68.0                            | 65.3                              | 0.7<br>(7.1)                                   | -2.7<br>(7.1)                                   | 3.4<br>(7.0)                                     |
| Two Relationships Report |                                  |                                 |                                   |  |   |  |
| Optional                 | 46.9                             | 31.3                            | 36.7                              | 15.6*<br>(7.6)                                 | 5.4<br>(7.7)                                    | 10.2<br>(7.8)                                    |
| Mandatory                | 64.6                             | 61.9                            | 67.3                              | 2.7<br>(7.6)                                   | 5.4<br>(7.8)                                    | -2.7<br>(8.0)                                    |

**Notes:** \* $p \leq 0.05$ ; Effect size between 0.40 and 0.50 or approximately a "medium effect size" (Cohen 1988).

gram performed worse than those with the set diagram in all but one case (mandatory cardinality understanding with the two relationship report).

We provide a brief commentary on our results. First, the average accuracy with the set diagram was statistically higher than that with the ER diagram for optional cardinalities for all three dependent measures. Second, a comparison of accuracy between the set+ and ER diagrams showed no statistical differences but most cases weighed in favor of the set+ diagram. Third, in general, the set diagram was better than the set+ diagram in all cases. Considering the differences in understanding between the set+ and the ER diagrams was not statistically significant, and that the set diagram was better than the set+ diagram, our findings suggest that strengthening the predominance of cardinality constraints (i.e., increasing the number of instances) in the set diagram (i.e., via the set+) does not further enhance understanding of the domain for our dependent measures. Instead, it seems that it could diminish understanding due to increased diagrammatic complexity.<sup>7</sup> Therefore, for our dependent measures, our results suggest that more concise extensional representations may have greater cognitive effectiveness (Larkin and Simon 1987) than

those that are less concise. Taking all experiment 1 findings into consideration, we further assess the advantage established by the set diagram compared to the ER diagram in experiment 2.

## Experiment 2

We designed experiment 2 to evaluate the differences in domain understanding with an ER diagram versus a set diagram for dependent measures different than the ones employed in experiment 1. The design of experiment 2 was similar to that of experiment 1.

### Experiment 2 Design

Similar to experiment 1, experiment 2 consisted of one between-subjects factor. The between-subjects factor, the type of representation, indicates that the participants received either a conceptual schema with cardinalities represented intensionally or extensionally. Similar to experiment 1, the dependent measures assessed understanding of the semantics of cardinality constraints; however, in experiment 2 we used different dependent measures than those used in experiment 1. A total of 100 participants were recruited from the same population of undergraduate business students as experiment

<sup>7</sup>By increasing the number of instances in the set+ diagram compared to the set diagram we increased the diagrammatic complexity of the set diagram (see Moody 2009).

1 and were randomly assigned to one of the two treatment groups: ER diagram versus set diagram. We assessed if we had successfully randomized our participants across the groups with respect to their backgrounds using  $F$ -tests, which suggested no significant differences between skill levels across the treatment groups. Also, a Pearson chi-square test ( $\chi^2 = 1.50$ ,  $df = 1$ ) and likelihood ratio test ( $\chi^2 = 1.50$ ,  $df = 1$ ) suggested that there was also no association between the representation treatment group assignment and participant gender.

## Experiment 2 Dependent Measures

This experiment consisted of two dependent measures. For each block of questions described below, half were related to optional and the other half to mandatory cardinality constraints. The order of mandatory and optional questions was randomized. First, elaborative reconstruction using memory recall (Bodart et al. 2001) consisted of six questions. The memory recall required participants to reconstruct the appropriate cardinality constraints without the use of the schema (see Appendix E). Participants responded to true/false questions to demonstrate their mental reconstruction of the cardinality constraints in their respective diagram.

The second dependent measure used was schema-based problem-solving, which has been widely employed in prior research (Shanks et al. 2008). The schema-based problem-solving required participants to envision scenarios within the domain that assessed their understanding of the cardinality-related semantics (see Appendix E). Participants could respond to the question by choosing one of three multiple-choice options: possible, not possible, or not sure. Our participants were also asked to provide a rationale for their answers.

## Experiment 2 Procedures

Due to the use of memory recall questions we had to provide ample opportunity for elaborative reconstruction. Therefore, experiment 2 required an alteration to the procedure employed in experiment 1. Participants first completed eight fill-in-the-blank questions as a warm-up task. The questions were designed to encourage traversal across the entire schema. Before completing the fill-in-the-blank questions, participants were first shown the correct answer to two fill-in-the-blank questions. The fill-in-the-blank questions were not scored, but instead only used to ensure that our participants spent ample time traversing the schema for memory pathways to be established. Next the participants completed the schema-based problem-solving questions, which were scored and

included in our analysis below. Then participants had the diagram taken from them before they completed the memory recall questions. The website could not advance until the lab administrator entered the appropriate password, thus ensuring that each participant's diagram was removed.

## Experiment 2 Analysis and Findings

We assessed two dependent measures: (1) memory recall; and (2) schema-based problem-solving to compare performance (percent correct) for the two groups. The mean accuracy for these questions for each treatment group can be found in Table 3 which indicates that for optional cardinalities, our participants performed significantly better with the set diagram than with the ER diagram. For questions that assessed understanding of mandatory cardinalities, we found that the differences between two groups were not statistically significant.

After each schema-based problem-solving question, we also asked participants to briefly explain the rationale for their answer. Evaluating the explanations required a judgement so we hired two independent coders who were not aware of the nature of this research. The coders were trained on cardinality constraints in conceptual modeling. The coders first watched the training videos provided to the participants and then one of the researchers provided additional training to the coders in using both the set and ER diagrams. In order to formalize the coding, the coders were provided with a coding scheme (shown in Appendix F). The coding scheme asked the coders to provide an overall score for each explanation. The overall score of the explanations were evaluated on the basis of (1) identification of appropriate elements from the schema and (2) application of appropriate logic when interpreting the semantics of cardinality constraints (see Appendix F).

The coders worked independently to code 20% of the data. They met to reconcile their different interpretations of the explanations and reported a Cohen's Kappa value of 0.93 (Cohen 1960) prior to reconciliation, and a Cohen's Kappa of 1.00 after reconciliation. Since the inter-rater agreement was sufficiently high, only one coder continued the rest of the coding.

The results of the coding can be seen in Table 4. Similar to our previous results with schema-based problem-solving, our findings suggest that participants using the set diagram better explained the semantics of optional cardinality constraints, but there was no significant difference in the quality of explanation of the semantics of mandatory cardinality constraints. This suggests that as compared to an ER diagram,

**Table 3. Detailed Results for Experiment 2**

| Task                         | Mean Accuracy (% Correct)     |                              | Mean Differences Set vs. ER (Std. Error) |
|------------------------------|-------------------------------|------------------------------|--|
|                              | Set Diagram<br>50 respondents | ER Diagram<br>50 respondents |  |
| Memory Recall                |                               |                              |  |
| Optional                     | 75.3                          | 51.0                         | 24.3**<br>(6.1)                          |
| Mandatory                    | 70.3                          | 71.3                         | 1.0<br>(5.0)                             |
| Schema-Based Problem Solving |                               |                              |  |
| Optional                     | 82.7                          | 68.7                         | 14.0*<br>(5.7)                           |
| Mandatory                    | 88.6                          | 87.3                         | 1.3<br>(4.5)                             |

Notes: \* $p \leq 0.05$ ; Effect size 0.49 or a "medium effect size." \*\* $p \leq 0.001$ ; effect size 0.79 or a "large effect size."

**Table 4. Coding Results for Schema-Based Problem-Solving Explanations**

| Task                                     | Quality of Explanation        |                              | Mean Differences Set vs. ER(Std. Error) |
|--|-------------------------------|------------------------------|---|
|  | Set Diagram<br>50 respondents | ER Diagram<br>50 respondents |   |
| Schema-Based Problem-Solving Explanation |                               |                              |   |
| Optional                                 | 5.9                           | 4.9                          | 1.00*<br>(0.48)                         |
| Mandatory                                | 6.7                           | 6.4                          | 0.30<br>(0.37)                          |

Notes: \* $p \leq 0.05$ ; effect size 0.39 or approximately a "medium effect size."

our participants better understood the semantics of optional cardinalities when using a set diagram.

Overall, our results are similar to our findings in experiment 1. Our participants performed better on dependent measures that assessed the understanding of the semantics of optional cardinalities with the Set diagram as opposed to the ER diagram. However, the differences in accuracy between the two groups were not statistically significant for dependent measures that assessed the understanding of the semantics of mandatory cardinalities.

## Discussion and Implications

This research set out to explore the effect of the type of encoding of semantics, intensional versus extensional, in a representation on domain understanding. Based on the arguments and presented findings, this research suggests that the tradition in IS research of exclusively focusing on the intensional encoding of domain semantics should be reexamined

and that the potential of extensional representations should be further examined.

We highlight our conclusions from this study. First, our research found that utilizing extensional representations are beneficial to understanding the semantics of the relationship cardinalities in conceptual schemas. We tested our proposition for two types of cardinalities, optional and mandatory. We found that, in general, the extensional representation outperformed the intensional counterpart for optional cardinalities. We did not find similar results for mandatory cardinalities. This finding is consistent with prior research which suggests that understanding optional cardinalities requires more cognitive effort than mandatory cardinalities (Bodart et al. 2001; Geminio and Wand 2005). Thus, there is a need for support mechanisms to overcome the cognitive difficulty of understanding optional cardinalities (Burton-Jones et al. 2012). As in prior research, we found that individuals performed significantly better on understanding mandatory than optional cardinalities for all tasks in our experiments. In experiment 1, the difference of the means of

mandatory understanding versus optional understanding was significant across all participants' performance pooled together in the set, set+, and ER diagram treatments (Understanding<sub>Mandatory</sub> = 70.52%; Understanding<sub>Optional</sub> = 48.83%;  $t = 8.03$ ,  $df = 146$ ,  $p \leq 0.001$ ). In experiment 2, the differences of the means was significant too (Understanding<sub>Mandatory</sub> = 79.42%; Understanding<sub>Optional</sub> = 69.42%;  $t = 3.62$ ,  $df = 99$ ,  $p \leq 0.001$ ).

We believe extensional representations are one type of cognitive support mechanism that can help in schema understanding. Our operationalization of explicating optionality-related semantics made the semantics more perceptive visually in the set diagram than in the ER diagram. In the ER diagram, the mandatory and optional cardinalities were distinguished with two different cardinality symbols. In the set diagram, mandatory cardinalities were encoded such that every entity instance was connected to a relationship instance, and optional cardinalities were encoded such that some entity instances did not connect to a relationship instance. This mixture of connectedness and disconnectness created a visual contrast that made the semantics of optional cardinalities more concrete, thus more coherent, for the respondents. Therefore, extensional encoding of optional cardinalities could have provided a cognitive support mechanism for understanding optional cardinalities. This explanation is further validated by research on concepts, which suggests that concrete- or instance-based representations are especially helpful for processing cognitively difficult information (Day and Goldstone 2012; Hahn and Chater 1998). Even in the case of elaborative reconstruction (i.e., our memory recall task), we believe extensional representations can serve as strong stimulation for activating pathways and can foster better recall.

Our use of multiple dependent measures helped address our proposition thoroughly, which is consistent with prior research. For example, Khatri et al. (2006) delineate dependent measures as read-to-recall (i.e., our memory recall) and read-to-do (i.e., our remaining dependent measures). Saghaei and Wand (2014) delineate dependent measures that require surface-level (i.e., our fill-in-the-blank, one-relationship report, two-relationship report, and memory recall) versus deep-level understanding (i.e., our schema-based problem-solving). We comprehensively employed several types of dependent measures that gauged surface- and deep-level understanding as well as both read-to-do and read-to-recall tasks.<sup>8</sup>

<sup>8</sup>We did not employ inferential reconstruction tasks as our goal was to measure actual understanding of a domain portrayed in a conceptual schema as opposed to inferred understanding of the domain.

## Applicability Check

To assess the external validity of our findings, we conducted an applicability check with practitioners following the guidance of Rosemann and Vessey (2008). To assess if the empirical benefits we found with respect to extensional representations might apply to practice, we conducted focus groups at the conclusion of our experiments. Furthermore, we took advantage of the focus groups to explore the potential applicability of extensional representations in other modeling topics.

Our focus groups took place at two different organizations during two sessions at each organization. Each session included participants from the respective organizations and two members of our research team. We sought to increase the diversity of organizations in order to assess the potential applicability of our findings with a broader audience. We thus sampled from organizations in two different industries, in life/health insurance and university domains. Henceforth, these organizations are referred to as InsuraCorp and EdU, respectively. Within each organization, we wanted to deliberate with individuals with strong conceptual modeling experience. As Table 5 shows, our focus group participants were indeed highly experienced.

Participants voluntarily joined an hour-long audio-recorded session at their respective organization. One research member was the focus group facilitator while the other took notes and asked follow-up questions to ensure no key ideas or concepts were missed. The first session at each organization began with brief introductions and participants completing a background questionnaire (see Table 5). A follow-up hour-long audio-recorded session was conducted at the participants' organization to clarify responses.

The focus group facilitator shared the session goals and instructions with the participants, discussed the motivation for our research, showed how we operationalized intensional and extensional representations, highlighted how we measured understanding of cardinalities, and then presented our overall results. During the presentation, extensional and intensional encoding of data semantics was framed as semantics explication with and without examples, respectively. This framing encapsulated the essence of our broader research program in a pragmatic fashion that would resonate with practice. After the presentation of the research results, the participants were led in a discussion using the following questions:

1. In general, should examples be used in (conceptual) modeling?
2. In general, why would examples help with understanding cardinalities?

**Table 5. Demographics of Focus Group Participants**

| Participant | Current IT Role    | Conceptual Modeling Experience | Models Created | Types of Models Created                               |
|-------------|--------------------|--------------------------------|----------------|---|
| InsuraCorp1 | Data Architect     | 30 years                       | 100+           | UML; MDM  |
| InsuraCorp2 | Data Architect     | 30 years                       | 100+           | ERD; dimensional; sequence; event; data flow          |
| InsuraCorp3 | Data Architect     | 25 years                       | 100+           | ERD; DFD; BPMN; TOGAF; UML                            |
| InsuraCorp4 | Data Architect     | 6.5 years                      | 50             | ERD; UML  |
| InsuraCorp5 | Data Architect     | 40 years                       | 100+           | ERD; UML  |
| EdU1        | Security Engineer  | 20 years                       | 50+            | BPMN; UML; ERD; TOGAF; use case                       |
| EdU2        | Software Architect | 15 years                       | 50+            | UML; ERD; BPMN; sequence; class; use case; flow chart |
| EdU3        | Data Architect     | 20 years                       | 100+           | ERD; UML; BPMN  |

3. Why would examples (Set Diagram) primarily help with understanding optional cardinalities?
4. How many examples do you think are necessary for understanding cardinalities?
5. What other potential applications of examples in modeling (beyond our research) exist?

While we motivated this research in the “Introduction” and “Background” sections with academic literature and textbooks, question 1 was asked to learn the need for this research from the perspective of practice. Question 2 was asked to assess the theoretical rationale provided in the proposition regarding examples helping with understanding cardinalities. Questions 3 and 4 were asked to understand the empirical findings of the research from the perspective of practice. Last, question 5 was asked to learn about possible extensions to our research. The participants were instructed to answer the questions candidly as there were no right answers, and to feel free to disagree with us or each other.

We summarize the main results of our findings in Table 6. First, the participants in the focus groups believed that examples get used for helping others understand conceptual models; however, the use of examples is informal or inconsistent. Our participants thought that examples improved modeling quality, enhanced communication of modeling, and should be mandatory for modeling. Second, the participants thought that examples help with understanding the semantics of cardinalities because of the concreteness effect. Additionally, they thought that the use of less technical jargon may have helped with domain understanding. Furthermore, the participants thought individuals employ examples to think through cardinality constraints, which in turn facilitates their ability to understand cardinalities. These reasons align with those we posited in the proposition of this research. Third, the participants thought the benefit of employing examples for

understanding the semantics of optional cardinalities was due to how the exception nature of optionality was better delineated visually in the set diagram. This provides another perspective on our discussion regarding extensional representations being a source of cognitive support. Fourth, with respect to the lack of empirical evidence for more examples in the set+ diagram, the focus group felt that the fewest number of examples that accurately convey semantics should be sufficient and that there is probably some inflection point regarding the benefit of the number of examples needed to better illustrate the meaning of cardinalities. This aligns with our empirical results in experiment 1 on the number of instances in an extensional representation. Finally, the participants in the focus groups identified several other applications where using examples in conceptual modeling could be beneficial in process models and infrastructure models. Beyond the scope of conceptual modeling, our participants mentioned using examples in test cases, software documentation, and business cases.<sup>9</sup> Based on our findings, we conclude that more research on extensional representations is warranted.

### Implications for Research

We highlight seven implications of this current work to research. First, while prior conceptual *schema creation* research suggests that individuals switch between concrete concepts (extension) to abstract concepts that describe a number of observations (intension) (Srinivasan and Te’eni 1995), to our knowledge this is the first paper to examine the effects of encoding domain semantics intensionally versus extensionally on *schema understanding*. Future research should compare the benefits and limitations of intensional

<sup>9</sup>Detailed quotes for the questions are available from the authors upon request.



**Table 6. Summary of Application Check Focus Group Results**

**Question 1: In general, should examples be used in (conceptual) modeling?**

*InsuraCorp*

- An example can be used (e.g., verbally, dry-erase board) to help people understand models.

*EdU*

- Examples are used informally and inconsistently in modeling.
- Examples in modeling help with communication with stakeholders and developers.
- Examples help improve modeling quality.
- Examples should be mandatory for modeling.

**Question 2: In general, why would examples help with understanding cardinalities?**

*InsuraCorp*

- Examples provide a different level of abstraction away from the class level and create a concreteness effect.
- Examples utilize domain language and less technical representations of cardinality constraints.
- Individuals instantiate to understand model relationships so examples facilitate their cognitive activity.

*EdU*

- Examples help individuals think through the nature of the relationships.
- Examples move the model from a concept to a more specific application to demonstrate relationships.
- Individuals picture or envision relationships in their thoughts or write them on paper to understand them.
- Examples in a diagram encode more information to communicate cardinality.

**Question 3: Why would examples (Set Diagram) primarily help with understanding optional cardinalities?**

*InsuraCorp*

- Optional cardinalities are an “exception rule” and examples better highlight exceptions.
- Visually seeing optionality with examples make its semantics apparent.
- The lack of lines (i.e., optionality) in the representation was a strong visual cue about the constraint.

*EdU*

- Optional cardinalities are outlier situations and examples bring them to life (attention).
- Individuals are “forced” to process or notice the optionality in the set diagram.

**Question 4: How many examples do you think are necessary for understanding cardinalities?**

*InsuraCorp*

- The smallest number that accurately convey the semantics.
- Every possible cardinality example that can exist.
- Enough to cover the known relationships.

*EdU*

- If optionality is present, you need at least one optional and one not optional.
- There is an inflection point in terms of cognitive tradeoff between too many and too few examples.
- Enough to allow individuals to quickly glance and notice optionality.

**Question 5: What other potential applications of examples in modeling (beyond our research) exist?**

*InsuraCorp*

- They could be used to illustrate the nuances of attribute values in conceptual data models (e.g., ER diagrams, data dictionaries).
- They could be used as in the movement of data via inputs and outputs of software functions such as data flow diagrams.
- They could be used for architecture diagrams such as (1) changes within systems and its impact on other systems; and (2) the replacement, retirement, and introduction of systems and their impact on other systems.

*EdU*

- Examples could be used in IT business cases (e.g., organizational change, organizational investment).
- Examples could be used in process models.
- Examples could be the basis for test cases.
- Examples could be used in software documentation shipped with software.
- Examples could be used in server/network Infrastructure physical layouts and diagrams.

representations and their extensional analog for the broader range of conceptual models such as process modeling, use case scenarios, programmer work specifications. Future research should also examine the extensional encoding of semantics of other types of relationships such as supertype–subtype and part–whole relationships (see Siau 2004). Finally, in the context of UML class diagrams, future research should also examine extensional encoding of methods that a class can execute.

Second, this research employed several types of dependent measures to assess the potential benefits/limitations of extensional and intensional representations. In doing so, we also conceptualized some tasks—specifically the report and the memory recall questions—differently than prior research. The report-based questions tested the ability of the respondents to understand implicit domain semantics. While our memory recall task was motivated by Bodart et al. (2001), we did not employ a multi-trial, free-recall, schema drawing task. We instead emphasized the recall of understanding the semantics of cardinality constraints. Future research should continue to develop and utilize dependent measures that are appropriate from both a research and a practical standpoint. Our focus was domain understanding; however, there are other dependent measures (e.g., information quality) that can be examined in future research on extensional representations and, more broadly, conceptual modeling.

Third, consistent with prior research in conceptual modeling (Bera et al. 2010; Burton-Jones et al. 2012; Dunn et al. 2011), the goal of this current research was to understand if there was an effect on domain understanding due to the type of representation used to encode semantics and we employed an experimental research method. We encourage research that employs other research methods. One benefit of another research method (e.g., process tracing) could be to provide deeper insights into how an individual processes semantics in intensional or extensional schemas used in this research, and even more broadly in other representations used in IS and its related disciplines. This might lead to being able to estimate the stage of information processing that led to the breakdown in performance with an intensional representation.

Fourth, prior research that proscribes the use of optional cardinalities on the basis of a violation of ontological soundness recommends using mandatory cardinalities exclusively (Bodart et al. 2001; Gemino and Wand 2005; Wand et al. 1995; Wand et al. 1999; Wand and Weber 1993).<sup>10</sup> While this

<sup>10</sup>Despite this proscription, conceptual schemas with optional associations still exist in research (Allen and March 2006, 2012) and practice (Elmasri and Navathe 2010; Geiger 2011; Hoffer et al. 2012). Recent NoSQL practice also embraces the use of optionality in representations (Hills 2016).

recommendation is theoretically sound, it also has the potential of increasing the complexity of a conceptual schema due to a larger number of the same type of constructs which in turn can negatively influence understanding of conceptual schemas. For example, while Burton-Jones et al. (2012) found that individuals could more accurately express the details of a conceptual schema when only mandatory cardinalities were used, performance with only mandatory cardinalities deteriorated as the diagrammatic complexity (see Moody (2009) in the sixth implication for research) of the conceptual schemas increased. We found that extensional representations can be a cognitive support mechanism that can help in domain understanding, in particular for optional cardinalities. Future research should continue to examine the benefit of extensional representations to support domain understanding. One idea could be to compare the effect of complexity on domain understanding from increasing the number of intensional constructs without optionality versus the complexity of using only extensional constructs while permitting optionality. Similar to cognitive science research, we also showed that the two types of representations—intensional and extensional—did not result in similar understanding (Hampton and Passanisi 2016), therefore future research should also examine ways to integrate the benefits of employing two types of encoding of domain semantics to increase overall domain understanding.

Fifth, in this study, we compared accuracy with only one type of representation—that is, intensional *or* extensional—for understanding. However, it is common for software development methodologies to employ multiple representations to holistically represent different aspects of a domain (Hungerford et al. 2004; Kim et al. 2000). With information distributed across several representations, those representations must facilitate an individual's ability to integrate information across different representations to form a coherent overarching understanding of the domain (Kim et al. 2000). To support integration, prior research has emphasized the importance of designing multiple related representations with (1) visual cues that indicate how an item in one representation is connected to the items in another representation (e.g., visual similarity of layouts) and (2) context information that facilitates an assessment of the semantics encoded in the representation (e.g., a separate representation that depicts how the other representations are interconnected at a high level) (Kim et al. 2000). Based on the performance on optional and mandatory cardinalities understanding, future research should examine the potential benefits of providing *both* an intensional representation and its extensional analog to an individual simultaneously.

Sixth, research is needed on how to create extensional representations from a design science perspective. We highlight four related implications.

- A new conceptual model for creating extensional representations might be needed. We created an extensional representation as an analog of an intensional representation. However, prior research has advocated creating extensional representations in lieu of intensional representations in contexts where data is generated outside the purview of an organization (e.g., citizen science and user-generated content) (Lukyanenko and Parsons 2012, 2013). Therefore, not only might a new conceptual model be necessary, but also guidance on when to create an extensional representation.
- Design guidance is also needed with respect to which instances should be used in the representation (Parsons 1996, 1997, 2000). There are several theories on concepts in cognitive psychology applicable to representations (for a review, see Lukyanenko and Samuel 2017) and these could have both cognitive implications for understanding extensional representations as well as design implications.
- Considering our earlier implication for research on multiple related representations, if the simultaneous availability of both intensional and extensional representations are found to be beneficial to domain understanding, it is also important to then understand *how* to integrate them. An approach for integrating multiple representations is important because prior research has found that individuals who are best able to use multiple representations are those who make round-trips between two representations instead of just random searches through given representations (Kim et al. 2000). We note that the challenge of how to integrate multiple representations together also applies to a context of multiple extensional representations.
- Moody (2009) suggests that the issue of diagrammatic complexity has an effect on the cognitive effectiveness of the diagram for a task. Complexity with intensional representations refers to the number of symbol instances in a given diagram (e.g., the number of entity types), but in the case of extensional representations it denotes the number of instances for both symbols and examples. Numerous symbol instances influence the complexity of the diagram (Moody 2002), and our comparisons of the set versus set+ diagrams illustrates that the number of example instances matter. Future research needs to advocate the apt number of symbol and example instances that can denote domain semantics while not significantly increasing the complexity of extensional representations. This also suggests the design of extensional representations will be impacted by the issue of its audience and purpose/task (see Samuel et al. 2015).

Our last implication notes that more research is needed to understand the effect of domain familiarity in conjunction with the use of extensional representations. Khatri et al. (2006) suggest that familiarity with a domain can also play a role in understanding. Future research should explore the boundaries to the benefits of extensional representations. Cognitive science work on schemas (Anderson 1980) might serve as a useful starting point. Cognitive schemas are large and complex units of knowledge that exist in a human mind and they contain much of what an individual knows about concepts. Perhaps in a familiar domain, individuals could benefit from extensional representations because the concepts fit into some schema of knowledge related to the domain that already exists in an individual's mind. On the other hand, in an unfamiliar domain, individuals might not have an existing schema of knowledge that they can employ to benefit from an extensional representation during understanding. Instead, individuals must first make sense of the concepts and develop their cognitive schema from the representation. Given that humans have limited cognitive resources (Miller 1956), future research should examine if individuals can simultaneously develop a cognitive schema in an unfamiliar domain from an extensional conceptual schema and still adequately perform a task.

### ***Implications for Practice***

In this research, we demonstrate the role that extensional encoding of semantics can play in understanding IS artifacts (representations) in a systems analysis and design context. Beyond our context, extensional representations can also play a role in information management. For example, in the context of master data management (MDM) and data integration, extensional representations can aid in denoting the semantics of attributes and cardinality constraints to facilitate an agreed-upon definition of data, thus facilitating management of non-transactional data, or master data, and integration of disparate data sources. As another example, big data analytics (Chen et al. 2012; LaValle et al. 2014; Minelli et al. 2012) seeks to generate insights by combining data from different systems, including externally generated data. In such a scenario, running automatic routines on data without fully understanding the data and their context can be detrimental (Provost and Fawcett 2013). Extensional representations can help individuals—for example, data scientists—who may not have been involved in the design of the data repositories (Browne and Ramesh 2002; Davenport and Patil 2012). Finally, research on citizen science has encouraged the use of extensional representations to improve information quality (Lukyanenko and Parsons 2012, 2013). In summary, extensional representations can play a role in better usage of data assets.

## Conclusion

While semantics of a representation can be encoded intensionally or extensionally, IS research to date has not evaluated the benefits and limitations of both types of representations on domain understanding. While prior research has embraced the challenge of improving the understandability of intensional representations to enhance domain understanding, it has overlooked the potential of extensional representations. Our research compared understanding of an intensional representation with its extensional analog. We used research on concepts and representations to conceptualize that extensional representations facilitate better understanding of the semantics of a domain more than intensional representations. Based on laboratory experiments, we found that understanding with an extensional representation was (1) at least as good as that with an intensional representation for mandatory cardinality constraints and (2) significantly better for optional cardinality constraints. This research suggests that in certain conditions, extensional representations can provide an avenue for improving understanding of the domain semantics.

## Acknowledgments

We thank the senior editor, associate editor, and anonymous reviewers for providing very helpful comments that improved the quality of the paper. We are grateful for the support that was provided by the Ivey Behavioural Research Lab at the University of Western Ontario, administered at the time by Lisa Bitacola, and the Kelley School of Business Behavioral Technology Lab at Indiana University, administered by Danielle Willibey.

## References

- Adams, L. T., Kasserman, J. E., Yearwood, A. A., Perfetto, G. A., Bransford, J. D., and Franks, J. J. 1988. "Memory Access: The Effects of Fact-Oriented Versus Problem-Oriented Acquisition," *Memory & Cognition* (16:2), pp. 167-175.
- Alexander, P. A. 1997. "Mapping the Multidimensional Nature of Domain Learning: The Interplay of Cognitive, Motivational, and Strategic Forces," in *Advances in Motivation and Achievement*, N. L. Maehr and P. R. Pintrich (eds.), Greenwich, CT: JAI Press, pp. 213-250.
- Allen, G., and March, S. 2006. "The Effects of State-Based and Event-Based Data Representation on User Performance in Query Formulation Tasks," *MIS Quarterly* (30:2), pp. 269-290.
- Allen, G. N., and March, S. T. 2012. "A Research Note on Representing Part-Whole Relations in Conceptual Modeling," *MIS Quarterly* (36:3), pp. 945-964.
- Allen, S. W., and Brooks, L. R. 1991. "Specializing the Operation of an Explicit Rule," *Journal of Experimental Psychology: General* (120:1), pp. 3-19.
- Ambler, S. W. 2014a. "Personas: An Agile Introduction" (available online at <http://www.agilemodeling.com/artifacts/personas.htm>; retrieved June 2, 2017).
- Ambler, S. W. 2014b. "UML 2 Object Diagrams: An Agile Introduction" (available online at <http://www.agilemodeling.com/artifacts/objectDiagram.htm>; retrieved December 21, 2015).
- Anderson, J. R. 1980. *Cognitive Psychology and Its Implications*, New York: W. H. Freeman and Company.
- Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., and Trollip, S. R. 1976. "Instantiation of General Terms," *Journal of Verbal Learning and Verbal Behavior* (15:6), pp. 667-679.
- Atkinson, R. K., Derry, S. J., Renkl, A., and Wortham, D. 2000. "Learning from Examples: Instructional Principles from the Worked Examples Research," *Review of Educational Research* (70:2), pp. 181-214.
- Batra, D., Hoffler, J. A., and Bostrom, R. P. 1990. "Comparing Representations with Relational and EER Models," *Communications of the ACM* (33:2), pp. 126-139.
- Bera, P., Burton-Jones, A., and Wand, Y. 2011. "Guidelines for Designing Visual Ontologies to Support Knowledge Identification," *MIS Quarterly* (35:4), pp. 883-908.
- Bera, P., Burton-Jones, A., and Wand, Y. 2014. "Research Note—How Semantics and Pragmatics Interact in Understanding Conceptual Models," *Information Systems Research* (25:2), pp. 401-419.
- Bera, P., Krasnoperova, A., and Wand, Y. 2010. "Using Ontology Languages for Conceptual Modeling," *Journal of Database Management* (21:1), pp. 1-28.
- Beveridge, M., and Parkins, E. 1987. "Visual Representation in Analogical Problem Solving," *Memory & Cognition* (15:3), pp. 230-237.
- Bodart, F., Patel, A., Sim, M., and Weber, R. 2001. "Should Optional Properties Be Used in Conceptual Modeling? A Theory and Three Empirical Tests," *Information Systems Research* (12:4), pp. 384-405.
- Bowen, P. L., O'Farrell, R. A., and Rohde, F. H. 2009. "An Empirical Investigation of End-User Query Development: The Effects of Improved Model Expressiveness Versus Complexity," *Information Systems Research* (20:4), pp. 565-584.
- Browne, G. J., and Ramesh, V. 2002. "Improving Information Requirements Determination: A Cognitive Perspective," *Information & Management* (39:8), pp. 625-645.
- Bunge, M. A. 1977. *Ontology 1: Furniture of the World*, Boston: Reidel Publishing Company.
- Burton-Jones, A., Clarke, R., Lazarenko, K., and Weber, R. 2012. "Is Use of Optional Attributes and Associations in Conceptual Modeling Always Problematic? Theory and Empirical Tests," in *Proceedings of the 33rd International Conference on Information Systems*, Orlando, FL, pp. 1-16.
- Burton-Jones, A., and Meso, P. N. 2006. "Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis," *Information Systems Research* (17:1), pp. 38-60.
- Burton-Jones, A., and Meso, P. 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), pp. 748-802.

- Burton-Jones, A., Wand, Y., and Weber, R. 2009. "Guidelines for Empirical Evaluations of Conceptual Modeling Grammars," *Journal of the Association for Information Systems* (10:6), pp. 495-532.
- Chen, P. P.-S. 1976. "The Entity-Relationship Model: Toward a Unified View of Data," *ACM Transactions on Database Systems* (1:1), pp. 9-36.
- Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20:1), pp. 37-46.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2<sup>nd</sup> ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins, A. M., and Quillian, M. R. 1969. "Retrieval Time from Semantic Memory," *Journal of Verbal Learning and Verbal Behavior* (8:2), pp. 240-247.
- Cooper, A. 2004. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*, Carmel, IN: Sams Publishing.
- Currim, F., and Ram, S. 2012. "Modeling Spatial and Temporal Set-Based Constraints During Conceptual Database Design," *Information Systems Research* (23:1), pp. 109-128.
- Davenport, T. H., and Patil, D. 2012. "Data Scientist," *Harvard Business Review* (90), pp. 70-76.
- Davies, I., Green, P., Rosemann, M., Indulska, M., and Gallo, S. 2006. "How Do Practitioners Use Conceptual Modeling in Practice?," *Data & Knowledge Engineering* (58:3), pp. 358-380.
- Day, S. B., and Goldstone, R. L. 2012. "The Import of Knowledge Export: Connecting Findings and Theories of Transfer of Learning," *Educational Psychologist* (47:3), pp. 153-176.
- Dennis, A., Wixom, B. H., and Tegarden, D. 2012. *Systems Analysis Design, UML Version 2.0: An Object Oriented Approach* (4<sup>th</sup> ed.), Hoboken, NJ: John Wiley & Sons.
- Dunn, C. L., Gerard, G. J., and Grabski, S. V. 2011. "Diagrammatic Attention Management and the Effect of Conceptual Model Structure on Cardinality Validation," *Journal of the Association for Information Systems* (12:8), pp. 585-605.
- Elmasri, R., and Navathe, S. 2010. *Fundamentals of Database Systems* (6<sup>th</sup> ed.), Boston: Addison-Wesley.
- Fettke, P. 2009. "How Conceptual Modeling Is Used," *Communications of the Association for Information Systems* (25:1), pp. 571-592.
- Frisch, J. C. 1969. *Extension and Comprehension in Logic*, New York: Philosophical Library.
- Geiger, J. G. 2011. "How to Validate the Data Model," *Information Management* (21:2), pp. 34-35.
- Gemino, A., and Wand, Y. 2005. "Complexity and Clarity in Conceptual Modeling: Comparison of Mandatory and Optional Properties," *Data & Knowledge Engineering* (55:3), pp. 301-326.
- Gillenson, M. L. 2012. *Fundamentals of Database Management Systems*, Hoboken, NJ: John Wiley & Sons.
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., and Kaschak, M. P. 2004. "Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension," *Journal of Educational Psychology* (96:3), pp. 424-436.
- Goldstone, R. L., and Medin, D. L. 1994. "Time Course of Comparison," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (20:1), pp. 29-50.
- Hahn, U., and Chater, N. 1998. "Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable?," *Cognition* (65:2), pp. 197-230.
- Hampton, J. A., and Passanisi, A. 2016. "When Intensions Do Not Map onto Extensions: Individual Differences in Conceptualization," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (42:4), pp. 505-523.
- Hills, T. 2016. *NoSQL and SQL Data Modeling*, Basking Ridge, NJ: Technics Publications.
- Hoffer, J. A., Ramesh, V., and Topi, H. 2012. *Modern Database Management* (11<sup>th</sup> ed.), Boston: Prentice Hall.
- Hoffer, J. A., Venkataraman, R., and Topi, H. 2015. *Modern Database Management*, London: Pearson Education.
- Hungerford, B., Hevner, A., and Collins, R. 2004. "Reviewing Software Diagrams: A Cognitive Study," *IEEE Transactions on Software Engineering* (30:2), pp. 82-96.
- Juhn, S., and Naumann, J. D. 1985. "The Effectiveness of Data Representation Characteristics on User Validation," in *Proceedings of the 6<sup>th</sup> International Conference on Information Systems*, Indianapolis, Indiana, pp. 212-226.
- Kaminski, J. A., and Sloutsky, V. M. 2012. "Representation and Transfer of Abstract Mathematical Concepts in Adolescence and Young Adulthood," in *The Adolescent Brain: Learning, Reasoning, and Decision Making*, V. F. Reyna, S. B. Chapman, M. R. Dougherty, and J. Confrey (eds.), Washington, DC: American Psychological Association, pp. 67-94.
- Khatri, V., Vessey, I., Ramesh, V., Clay, P., and Park, S.-J. 2006. "Understanding Conceptual Schemas: Exploring the Role of Application and IS Domain Knowledge," *Information Systems Research* (17:1), pp. 81-99.
- Kifer, M., and Lausen, G. 1989. "F-Logic: A Higher-Order Language for Reasoning About Objects, Inheritance, and Scheme," in *Proceedings of the 1989 ACM SIGMOD International Conference*, pp. 134-146.
- Kim, J., Hahn, J., and Hahn, H. 2000. "How Do We Understand a System with (So) Many Diagrams? Cognitive Integration Processes in Diagrammatic Reasoning," *Information Systems Research* (11:3), pp. 284-303.
- Kim, Y.-G., and March, S. T. 1995. "Comparing Data Modeling Formalisms," *Communications of the ACM* (38:6), pp. 103-115.
- Koedinger, K. R., Alibali, M. W., and Nathan, M. J. 2008. "Trade-Offs between Grounded and Abstract Representations: Evidence from Algebra Problem Solving," *Cognitive Science* (32:2), pp. 366-397.
- Koedinger, K. R., and Nathan, M. J. 2004. "The Real Story Behind Story Problems: Effects of Representations on Quantitative Reasoning," *The Journal of the Learning Sciences* (13:2), pp. 129-164.
- Kuske, S., Gogolla, M., Kollmann, R., and Kreowski, H.-J. 2002. "An Integrated Semantics for UML Class, Object and State Diagrams Based on Graph Transformation," in *Integrated Formal Methods*, M. Butler, L. Petre, and K. Sere (eds.), Berlin: Springer, pp. 11-28.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, Chicago: University of Chicago Press.
- Larkin, J. H., and Simon, H. A. 1987. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words," *Cognitive Science* (11:1), pp. 65-100.
- Lave, J. 1988. *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life*, New York: Cambridge University Press.

- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., and Kruschwitz, N. 2014. "Big Analytics and the Path from Insights to Value," *Sloan Management Review* (52:2), pp. 21-31.
- Lecluse, C., Richard, P., and Velez, F. 1988. "O2, an Object-Oriented Data Model," in *Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data*, pp. 424-433.
- Liddle, S. W., Embley, D. W., and Woodfield, S. N. 1993. "Cardinality Constraints in Semantic Data Models," *Data & Knowledge Engineering* (11:3), pp. 235-270.
- Lockhart, R. S., Lamon, M., and Gick, M. L. 1988. "Conceptual Transfer Insimple Insight Problems," *Memory & Cognition* (16:1), pp. 36-44.
- Lukyanenko, R., and Parsons, J. 2012. "Conceptual Modeling Principles for Crowdsourcing," in *Proceedings of the 1st International Workshop on Multimodal Crowd Sensing*, Maui, Hawaii, pp. 3-6.
- Lukyanenko, R., and Parsons, J. 2013. "Is Traditional Conceptual Modeling Becoming Obsolete?," in *Conceptual Modeling*, W. Ng, V. Storey, and J. Trujillo (eds.), Berlin: Springer, pp. 61-73.
- Lukyanenko, R., and Samuel, B. M. 2017. "Are All Classes Created Equal? Increasing Precision of Conceptual Modeling Grammars," *ACM Transactions on Management Information Systems* (8:4), Article 14.
- Ma, J., and LeRouge, C. 2007. "Introducing User Profiles and Personas into Information Systems Development," in *Proceedings of the 2007 Americas Conference on Information Systems*, Keystone, CO, p. 237.
- Mark, L., and Roussopoulos, N. 1987. "Meta-Data Management," unpublished paper, University of Maryland, College Park, MD.
- Markovits, H., and Vachon, R. 1990. "Conditional Reasoning, Representation, and Level of Abstraction," *Developmental Psychology* (26:6), pp. 942-951.
- Mayer, R. E. 2001. *Multimedia Learning*, New York: Cambridge University Press.
- McNeil, N. M., Uttal, D. H., Jarvin, L., and Sternberg, R. J. 2009. "Should You Show Me the Money? Concrete Objects Both Hurt and Help Performance on Mathematics Problems," *Learning and Instruction* (19:2), pp. 171-184.
- Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review* (63:2), pp. 81-97.
- Moody, D. 2002. "Complexity Effects on End User Understanding of Data Models: An Experimental Comparison of Large Data Model Representation Methods," in *Proceedings of the 10th European Conference on Information Systems*, Gdansk, Poland.
- Moody, D. L. 2009. "The 'Physics' of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering," *Software Engineering, IEEE Transactions on* (35:6), pp. 756-779.
- Moreno, R., Ozogul, G., and Reisslein, M. 2011. "Teaching with Concrete and Abstract Visual Representations: Effects on Students' Problem Solving, Problem Representations, and Learning Perceptions," *Journal of Educational Psychology* (103:1), pp. 32-47.
- Motro, A. 1994. "Intensional Answers to Database Queries," *IEEE Transactions on Knowledge and Data Engineering* (6:3), pp. 444-454.
- Nosofsky, R. M. 1991. "Typicality in Logically Defined Categories: Exemplar-Similarity Versus Rule Instantiation," *Memory & Cognition* (19:2), pp. 131-150.
- Ohlsson, S., and Lehtinen, E. 1997. "Abstraction and the Acquisition of Complex Ideas," *International Journal of Educational Research* (27:1), pp. 37-48.
- Olivé, A. 2007. "Cardinality Constraints," Chapter 4 in *Conceptual Modeling of Information Systems*. Berlin: Springer, pp. 83-102.
- OMG. 2015. "Semantics of Business Vocabulary and Business Rules, Version 1.3" (available online at <http://www.omg.org/spec/SBVR/1.3/index.htm>).
- Paivio, A. 2007. *Mind and Its Evolution: A Dual Coding Theoretical Approach*, Mahwah, NJ: Lawrence Erlbaum Associates.
- Parsons, J. 1996. "An Information Model Based on Classification Theory," *Management Science* (42:10), pp. 1437-1453.
- Parsons, J. 2002. "Effects of Local Versus Global Schema Diagrams on Verification and Communication in Conceptual Data Modeling," *Journal of Management Information Systems* (19:3), pp. 155-183.
- 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), pp. 441-462.
- Parsons, J., and Wand, Y. 1997. "Choosing Classes in Conceptual Modeling," *Communications of the ACM* (40:6), pp. 63-69.
- Parsons, J., and Wand, Y. 2000. "Emancipating Instances from the Tyranny of Classes in Information Modeling," *ACM Transactions on Database Systems* (25:2), pp. 228-268.
- Parsons, J., and Wand, Y. 2008. "Using Cognitive Principles to Guide Classification in Information Systems Modeling," *MIS Quarterly* (32:4), pp. 839-868.
- Pilone, D., and Pitman, N. 2005. *UML 2.0 in a Nutshell* (2nd ed.), Sebastopol, CA: O'Reilly Media Inc.
- Posner, M. I., and Keele, S. W. 1970. "Retention of Abstract Ideas," *Journal of Experimental Psychology* (83), pp. 304-308.
- Provost, F., and Fawcett, T. 2013. *Data Science for Business: What you Need to Know About Data Mining and Data-Analytic Thinking*, Sebastopol, CA: O'Reilly Media, Inc.
- Ram, S., and Khatri, V. 2005. "A Comprehensive Framework for Modeling Set-Based Business Rules During Conceptual Database Design," *Information Systems* (30:2), pp. 89-118.
- Recker, J., Rosemann, M., Green, P., and Indulska, M. 2011. "Do Ontological Deficiencies in Modeling Grammars Matter?," *MIS Quarterly* (35:1), pp. 1-24.
- Rosca, D., Greenspan, S., and Wild, C. 2002. "Enterprise Modeling and Decision-Support for Automating the Business Rules Lifecycle," *Automated Software Engineering* (9:4), pp. 361-404.
- Rosch, E. H. 1973. "Natural Categories," *Cognitive Psychology* (4:3), pp. 328-350.
- Rosch, E. 1978. "Principles of Categorization," in *Cognition and Categorization*, E. Rosch and B. Lloyd (eds.), Hillsdale, NJ: Erlbaum, pp. 27-48.
- Rosemann, M., and Vessey, I. 2008. "Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks," *MIS Quarterly* (32:1), pp. 1-22.
- Roth, W. M., and Hwang, S. W. 2006. "Does Mathematical Learning Occur in Going from Concrete to Abstract or in Going from Abstract to Concrete?," *The Journal of Mathematical Behavior* (25:4), pp. 334-344.

- Rundensteiner, E. A. 1992. "Set Operations in Object-Based Data Models," *IEEE Transactions on Knowledge and Data Engineering* (4), pp. 382-398.
- Rundensteiner, E. A., Bic, L., Gilbert, J. P., and Yin, M. L. 1994. "Set Restrictions for Semantic Groupings," *IEEE Transactions on Knowledge and Data Engineering* (6:2), pp. 193-204.
- Saghafi, A., and Wand, Y. 2014. "Do Ontological Guidelines Improve Understandability of Conceptual Models? A Meta-Analysis of Empirical Work," in *Proceedings of the 47<sup>th</sup> Hawaii International Conference on System Sciences*, pp. 4609-4618.
- Samuel, B. M., Watkins III, L. A., Ehle, A., and Khatri, V. 2015. "Customizing the Representation Capabilities of Process Models: Understanding the Effects of Perceived Modeling Impediments," *IEEE Transactions on Software Engineering* (41:1), pp. 19-39.
- Savnik, I., Tari, Z., and Mohorič, T. 1999. "QAL: A Query Algebra of Complex Objects," *Data & Knowledge Engineering* (30:1), pp. 57-94.
- Shanks, G., Nuredini, J., Tobin, D., and Weber, R. 2003. "Representing Things and Properties in Conceptual Modeling: Understanding the Impact of Task Type," in *Proceedings of the 24<sup>th</sup> International Conference on Information Systems*, Seattle, pp. 909-913.
- Shanks, G., Tansley, E., Nuredini, J., Tobin, D., and Weber, R. 2008. "Representing Part-Whole Relations in Conceptual Modeling: An Empirical Evaluation," *MIS Quarterly* (32:3), pp. 553-573.
- Siau, K. 2004. "Relationship Construct in Modeling Information Systems: Identifying Relationships Based on Relation Element Theory," *Journal of Database Management* (15:3), pp. i-v.
- Silberschatz, A., Korth, H. F., and Sudarshan, S. 2011. *Database System Concepts*, New York: McGraw-Hill.
- Srinivasan, A., and Te'eni, D. 1995. "Modeling as Constrained Problem Solving: An Empirical Study of the Data Modeling Process," *Management Science* (41:3), pp. 419-434.
- Tashakkori, A., and Teddlie, C. 1998. *Mixed Methodology: Combining Qualitative and Quantitative Approaches*, Thousand Oaks, CA: Sage.
- Thalheim, B. 1992. "Fundamentals of Cardinality Constraints," in *Entity-Relationship Approach—ER '92*, G. Pernul and A. M. Tjoa (eds.), Berlin: Springer, pp. 7-23.
- Umanath, N. S., and Scamell, R. W. 2015. *Data Modeling and Database Design*, Boston: Cengage Learning.
- Wand, Y., Monarchi, D., Parsons, J., and Woo, C. 1995. "Theoretical Foundations for Conceptual Modeling in Information Systems Development," *Decision Support Systems* (15), pp. 285-304.
- Wand, Y., Storey, V. C., and Weber, R. 1999. "An Ontological Analysis of the Relationship Construct in Conceptual Modeling," *ACM Transactions on Database Systems* (24:4), pp. 494-528.
- Wand, Y., and Weber, R. 1993. "On the Ontological Expressiveness of Information Systems Analysis and Design Grammars," *Journal of Information Systems* (3), pp. 217-237.
- Whitney, C., Grossman, M., and Kircher, T. T. J. 2009. "The Influence of Multiple Primes on Bottom-Up and Top-Down Regulation During Meaning Retrieval: Evidence for 2 Distinct Neural Networks," *Cerebral Cortex* (19:11), pp. 2548-2560.
- Zhou, B., and Mao, Y. 2010. "Four Semantic Layers of Common Nouns," *Synthese* (175:1), pp. 47-68.

## About the Authors

**Binny M. Samuel** is an Assistant Professor in the Operations, Business Analytics, and Information Systems Department at the University of Cincinnati's Lindner College of Business. He earned his Ph.D. from the Kelley School of Business at Indiana University. He also holds a Bachelor of Science in Business and an M.B.A, with concentrations in Accounting and Information Systems. Prior to his doctoral education, he worked in IT roles at Ford Motor Company and at Indiana University. Prior to joining the University of Cincinnati, he worked at the Ivey Business School at the University of Western Ontario. His work has been published in *ACM Transactions on Management Information Systems*, *AIS Transactions on Replication Research*, *European Journal of Information Systems*, *IEEE Transactions on Software Engineering*, *Information & Management*, and *Journal of Information Technology*.

**Vijay Khatri** is a Professor in the Operations and Decision Technologies Department at the Kelley School of Business. He holds a B.E. from Malaviya National Institute of Technology, a management degree from the University of Bombay, and a Ph.D. from the University of Arizona. His research centers on issues related to data semantics, semiotics and conceptual database design, temporal databases, and data governance. Specifically, his research involves developing conceptual design techniques for management of data, especially for applications that need to organize data based on time and space. He has published articles in journals such as *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Software Engineering*, *Annals of Mathematics and Artificial Intelligence*, *Information Systems Research*, *Journal of Management Information Systems*, and *Communications of the ACM*. Previously, he has worked with Infosys Technologies and IBM Consulting Group.

**Ramesh Venkataraman**, is Associate Dean for Information and Instructional Technologies, Professor of Information Systems and John R. Gibbs Professor at Indiana University's Kelley School of Business. He is also chairman of the Kelley Direct Online MBA and MS programs. He has published over 25 papers in leading journals, such as *Information Systems Research*, *MIS Quarterly*, *ACM Transactions on Information Systems*, *Communications of the ACM*, and *Journal of Management Information Systems*. Ramesh is a co-author on one of the leading database books, *Modern Database Management* (13<sup>th</sup> ed.), with Jeff Hoffer and Heikki Topi.

## EXPLORING THE EFFECTS OF EXTENSIONAL VERSUS INTENSIONAL REPRESENTATIONS ON DOMAIN UNDERSTANDING

**Binny M. Samuel**

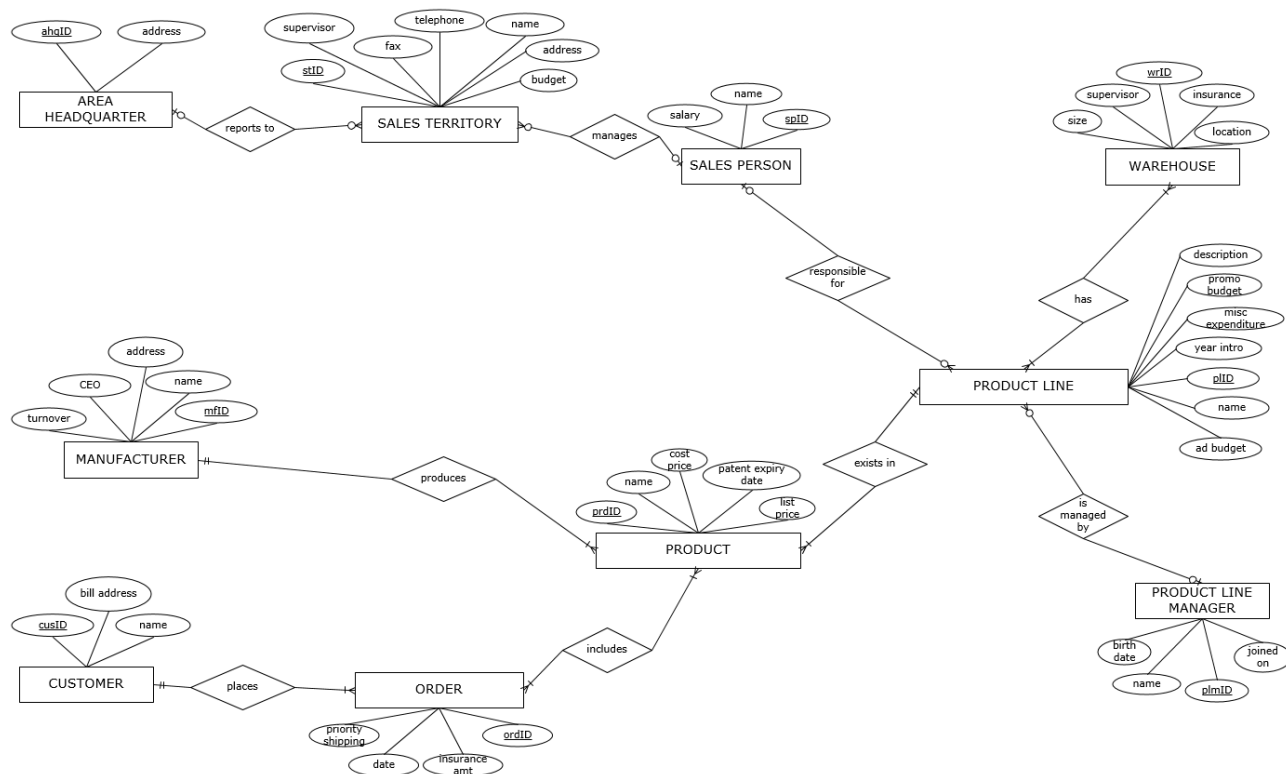
Carl H. Lindner College of Business, University of Cincinnati, PO Box 210130,  
Cincinnati, OH 45221-0130 U.S.A. {samuelby@uc.edu}

**Vijay Khatri and V. Ramesh**

Kelley School of Business, Indiana University, 1309 E. 10<sup>th</sup> Street,  
Bloomington, IN 47405 U.S.A. {vkhatr@indiana.edu} {venkat@indiana.edu}

### Appendix A

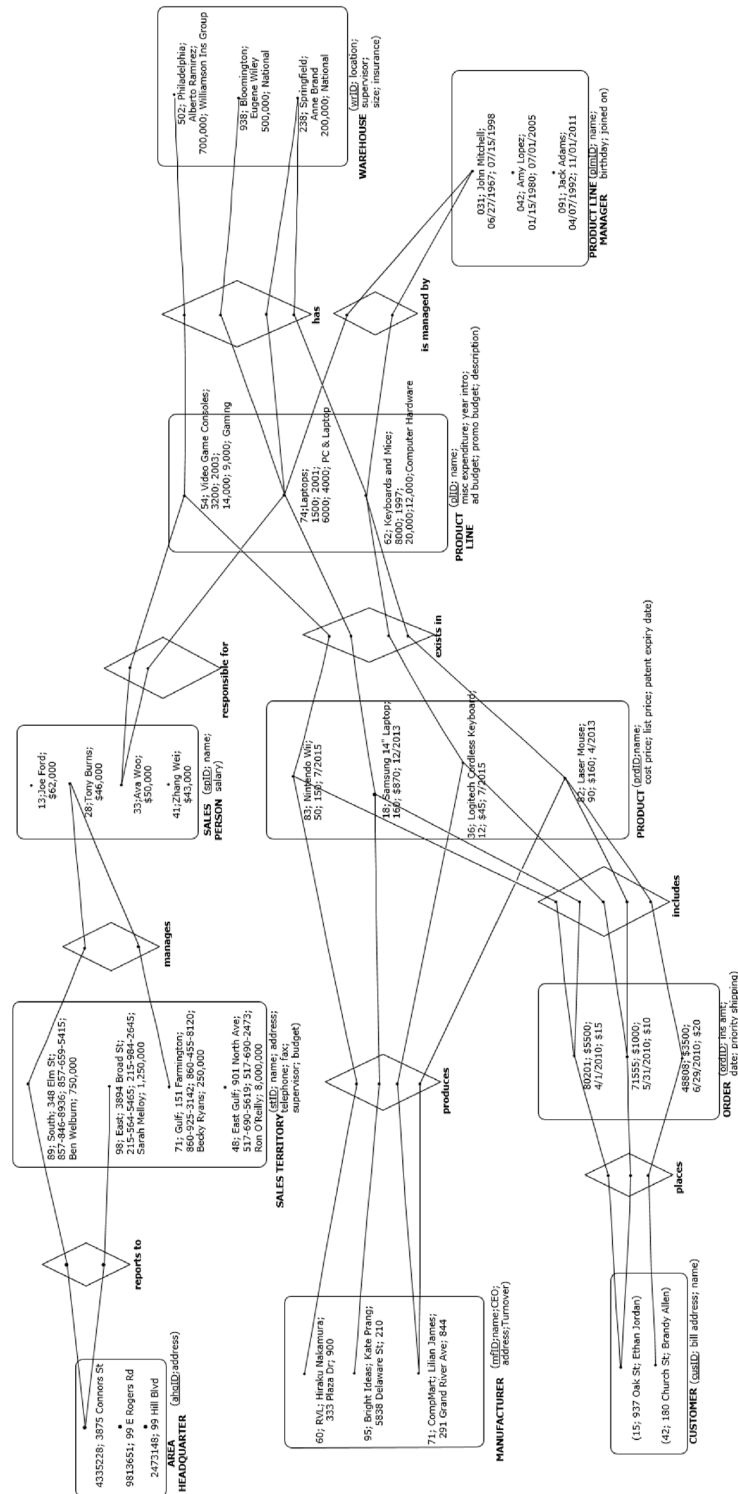
#### ER Diagram for Sales Domain





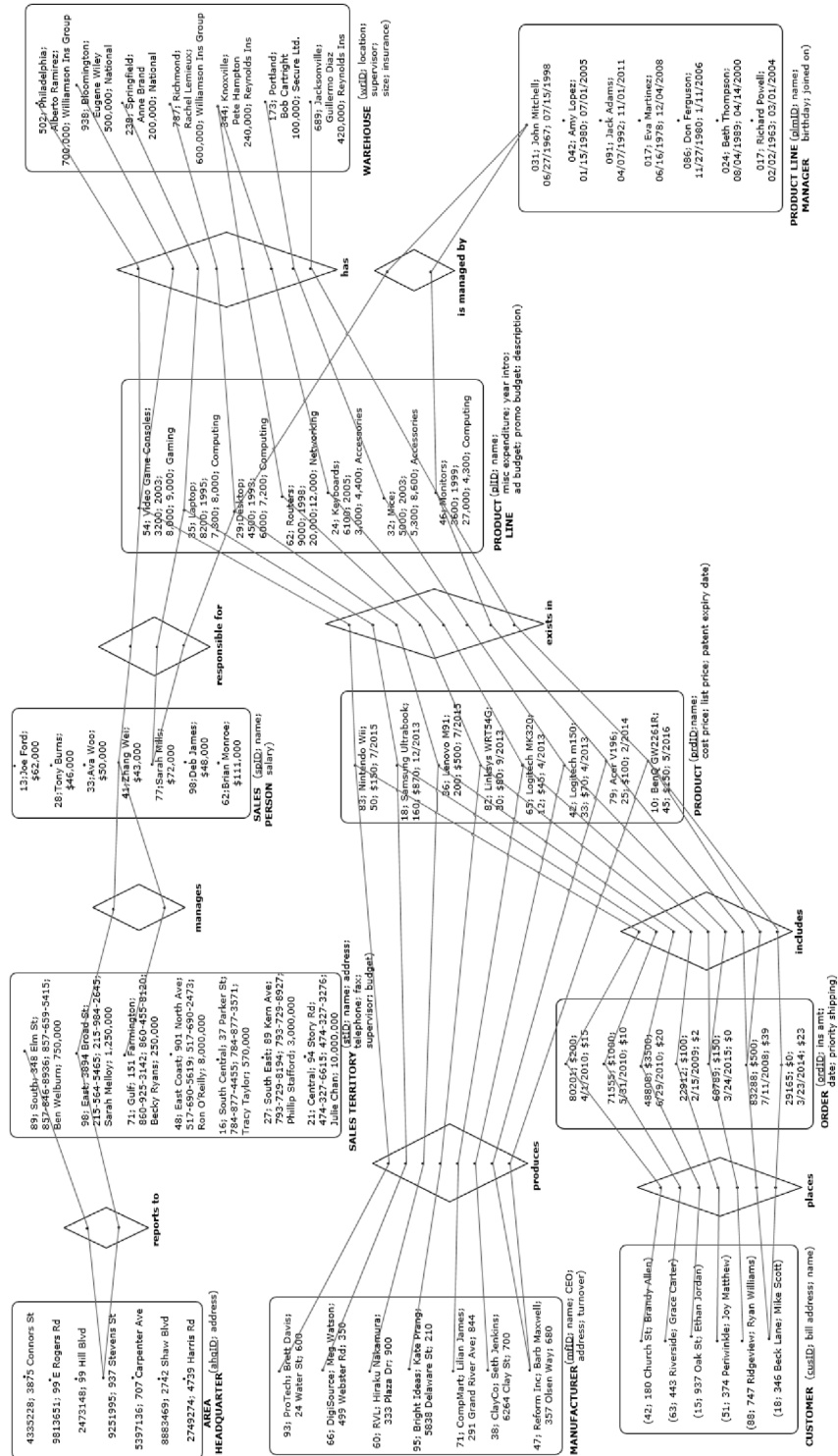
# Appendix B

## Set Diagram for Sales Domain



# Appendix C

## Set+ Diagram for Sales Domain



## Appendix D

### Pre-experiment Questions to Assess Background of Participants

| I am skilled with the following:                               | Strongly Disagree | Disagree | Slightly Disagree | Neither Agree nor Disagree | Slightly Agree | Agree | Strongly Agree |
|--|-------------------|----------|-------------------|----------------------------|----------------|-------|----------------|
| Using conceptual models (for example, ER Diagram)              |                   |          |                   |                            |                |       |                |
| Writing SQL (Structured Query Language)                        |                   |          |                   |                            |                |       |                |
| Writing computer programming language code (for example, Java) |                   |          |                   |                            |                |       |                |
| Sales management, order processing, and inventory management   |                   |          |                   |                            |                |       |                |

# Appendix E

## Task Questions

### Fill-in-the-Blank Questions (Experiment 1)

For this task, the respondents had to identify the minimum cardinalities from the schema provided to them. For example, for question 3 (mandatory cardinality), the answer was “one” because every order instance is associated with at least one customer instance. Those who received the ER diagram based that answer on the syntax of cardinality symbols. On the other hand, those who received the set diagram had cardinalities encoded with instances, which emphasized mandatory cardinality using three instances (80201, 71555, and 48808) wherein all three instances participated in the relationship places. Those with the ER diagram did not get the benefit of a ready-made example and had to create the extension of the schema for the generic instance in the question (i.e., “an order”), which required additional processing.

### Optional Cardinality

1. An area headquarter has reporting to it a minimum of \_\_\_\_\_ sales territory(ies).
2. A sales territory is managed by a minimum of \_\_\_\_\_ sales person(s).
3. A sales person is responsible for a minimum of \_\_\_\_\_ product line(s).

### Mandatory Cardinality

1. A warehouse has in it a minimum of \_\_\_\_\_ product line(s).
2. A product is included in a minimum of \_\_\_\_\_ order(s).
3. An order is placed by a minimum of \_\_\_\_\_ customer(s).

### Report Questions: One Relationship (Experiment 1)

Given the cardinality specification in the schema, the respondents needed to verify if the instances indicated in the report were valid. This task relied on reports, or queries, which are a way of making data semantics (in the schema) explicit. For example, for question 3 (mandatory cardinality), the answer was “Only Row # 2” (b). Given that the relationship between PRODUCT and PRODUCT LINE was mandatory, each product has to have a product line associated with it. This task seeks to gauge the ability to understand the semantics that are explicated in the extension of the relationship “exists in.” Those who received the set diagram received ready-made examples which they could use to connect with the given report and test out their understanding of the report. Those who received the ER diagram had to create the extension of the schema and map that to the report that was given to them, which required additional processing.

### Optional Cardinality

1. A report has been created about SALES TERRITORY and SALES PERSON (see the diagram). Below is an excerpt of the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* SALES TERRITORY and any associated SALES PERSON. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | SALES TERRITORY<br>stID | SALES TERRITORY<br>name | SALES PERSON<br>spID | SALES PERSON<br>salary |
|-------|-------------------------|-------------------------|----------------------|------------------------|
| 1     | stID1                   | name1                   | spID2                | salary2                |
| 2     | stID3                   | name3                   |                      |                        |
| 3     | stID6                   | name4                   | spID5                | salary5                |

- (a) Only Row #1
- (b) Only Row #2

- (c) Only Row #1 and Row #2  
 (d) Only Row #1 and Row #3  
 (e) All of the rows are possible
2. A report has been created about SALES PERSON and PRODUCT LINE (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* SALES PERSON and any associated PRODUCT LINE. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | SALES PERSON<br>spID | SALES PERSON<br>name | PRODUCT LINE<br>plID | PRODUCT LINE<br>name |
|-------|----------------------|----------------------|----------------------|----------------------|
| 1     | spID1                | name1                | plID4                | name4                |
| 2     | spID5                | name5                |                      |                      |
| 3     | spID9                | name9                |                      |                      |

- (a) Only Row #1  
 (b) Only Row #2  
 (c) Only Row #1 and Row #2  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible
3. A report has been created about PRODUCT LINE and PRODUCT LINE MANAGER (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* PRODUCT LINE and any associated PRODUCT LINE MANAGER. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | PRODUCT LINE<br>plID | PRODUCT LINE ad<br>budget | PRODUCT LINE<br>MANAGER plmID | PRODUCT LINE<br>MANAGER joined on |
|-------|----------------------|---------------------------|-------------------------------|-----------------------------------|
| 1     | plID2                | ad budget2                |                               |                                   |
| 2     | plID4                | ad budget4                | plmID3                        | joined on3                        |
| 3     | plID1                | ad budget1                |                               |                                   |

- (a) Only Row #2  
 (b) Only Row #3  
 (c) Only Row #1 and Row #2  
 (d) Only Row #1 and Row #3  
 (e) All of the rows are possible

### Mandatory Cardinality

1. A report has been created about CUSTOMER and ORDER (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* CUSTOMER and any associated ORDER. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | CUSTOMER cusID | CUSTOMER name | ORDER ordID | ORDER date |
|-------|----------------|---------------|-------------|------------|
| 1     | cusID5         | name5         | ordID2      | date2      |
| 2     | cusID7         | name7         | ordID1      | date1      |
| 3     | cusID3         | name3         |             |            |

- (a) Only Row #1  
 (b) Only Row #2  
 (c) Only Row #1 and Row #2  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible

2. A report has been created about WAREHOUSE and PRODUCT LINE (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* WAREHOUSE and any associated PRODUCT LINE. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | WAREHOUSE wrID | WAREHOUSE size | PRODUCT LINE plID | PRODUCT LINE year intro |
|-------|----------------|----------------|-------------------|-------------------------|
| 1     | wrID1          | size1          | plID4             | year intro4             |
| 2     | wrID2          | size2          |                   |                         |
| 3     | wrID9          | size9          |                   |                         |

- (a) Only Row #1  
 (b) Only Row #3  
 (c) Only Row #1 and Row #2  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible
3. A report has been created about PRODUCT and PRODUCT LINE (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* PRODUCT and any associated PRODUCT LINE. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | PRODUCT prdID | PRODUCT list price | PRODUCT LINE plID | PRODUCT LINE ad budget |
|-------|---------------|--------------------|-------------------|------------------------|
| 1     | prdID4        | list price4        |                   |                        |
| 2     | prdID1        | list price1        | plID3             | ad budget3             |
| 3     | prdID2        | list price2        |                   |                        |

- (a) Only Row #3  
 (b) Only Row #2  
 (c) Only Row #1 and Row #2  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible

### Report Questions: Two Relationships (Experiment 1)

These tasks were similar to those presented in the previous section, except that they required traversal of two relationships. Those individuals given a set diagram could better understand the semantics across two relationships using instantiated relationships (see the explanation for the previous section). For example, for question 1 (optional cardinality) the answer was “All of the rows are possible” (e). Given that the relationships between AREA HEADQUARTER, SALES TERRITORY and SALES PERSON were each optional in the schema, the respondents could associate the optionality in the set diagram with that in the report directly. Those with the ER diagram needed create an extension of the schema and then map that to the report given to them, which required additional processing.

### Optional Cardinality

1. A report has been created about AREA HEADQUARTER and SALES PERSON (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* AREA HEADQUARTER and any associated SALES PERSON. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | AREA HEADQUARTER<br>ahqID | AREA HEADQUARTER<br>address | SALES PERSON<br>spID | SALES PERSON<br>salary |
|-------|---------------------------|-----------------------------|----------------------|------------------------|
| 1     | ahqID1                    | address1                    | spID3                | salary3                |
| 2     | ahqID2                    | address2                    |                      |                        |
| 3     | ahqID5                    | address5                    |                      |                        |

- (a) Only Row #1  
 (b) Only Row #2  
 (c) Only Row #3  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible
2. A report has been created about SALES TERRITORY and PRODUCT LINE (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* SALES TERRITORY and any associated PRODUCT LINE. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | SALES TERRITORY<br>stID | SALES TERRITORY<br>budget | PRODUCT LINE<br>plID | PRODUCT LINE ad<br>budget |
|-------|-------------------------|---------------------------|----------------------|---------------------------|
| 1     | stID5                   | budget5                   | pdID7                | ad budget7                |
| 2     | stID2                   | budget2                   |                      |                           |
| 3     | stID4                   | budget4                   | plID1                | ad budget1                |

- (a) Only Row #3  
 (b) Only Row #2  
 (c) Only Row #1 and Row #3  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible
3. A report has been created about PRODUCT LINE MANAGER and SALES PERSON (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* PRODUCT LINE MANAGER and any associated SALES PERSON. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | PRODUCT LINE<br>MANAGER plmID | PRODUCT LINE<br>MANAGER name | SALES PERSON<br>spID | SALES PERSON<br>name |
|-------|-------------------------------|------------------------------|----------------------|----------------------|
| 1     | plmID3                        | name3                        |                      |                      |
| 2     | plmID4                        | name4                        | spID1                | name1                |
| 3     | plmID2                        | name2                        |                      |                      |

- (a) Only Row #1  
 (b) Only Row #2  
 (c) Only Row #1 and Row #3  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible

### Mandatory Cardinality

1. A report has been created about CUSTOMER and PRODUCT (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* CUSTOMER and any associated PRODUCT. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | CUSTOMER cusID | CUSTOMER name | PRODUCT prdID | PRODUCT list price |
|-------|----------------|---------------|---------------|--------------------|
| 1     | cusID1         | name1         | prdID3        | list price3        |
| 2     | cusID2         | name2         |               |                    |
| 3     | cusID8         | name8         | prdID6        | list price6        |

- (a) Only Row #1  
 (b) Only Row #2  
 (c) Only Row #1 and Row #3  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible
2. A report has been created about PRODUCT LINE and MANUFACTURER (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* PRODUCT LINE and any associated MANUFACTURER. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | PRODUCT LINE plID | PRODUCT LINE description | MANUFACTURER mflID | MANUFACTURER turnover |
|-------|-------------------|--------------------------|--------------------|-----------------------|
| 1     | plID1             | description1             | mflID4             | turnover4             |
| 2     | plID3             | description3             |                    |                       |
| 3     | plID7             | description7             |                    |                       |

- (a) Only Row #1  
 (b) Only Row #2  
 (c) Only Row #1 and Row #3  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible
3. A report has been created about PRODUCT and WAREHOUSE (see the diagram). Below is the report template which contains column headers and placeholders for actual data that is modeled by the diagram provided to you. The actual report will contain many more rows, displaying *every* PRODUCT and any associated WAREHOUSE. Based on your understanding of the cardinality rules conveyed by the diagram, which of the rows in the excerpt are possible?

| Row # | PRODUCT prdID | PRODUCT cost price | WAREHOUSE wrID | WAREHOUSE supervisor |
|-------|---------------|--------------------|----------------|----------------------|
| 1     | prdID6        | cost price6        |                |                      |
| 2     | prdID2        | cost price2        | wrID5          | supervisor5          |
| 3     | prdID1        | cost price1        | wrID4          | supervisor4          |

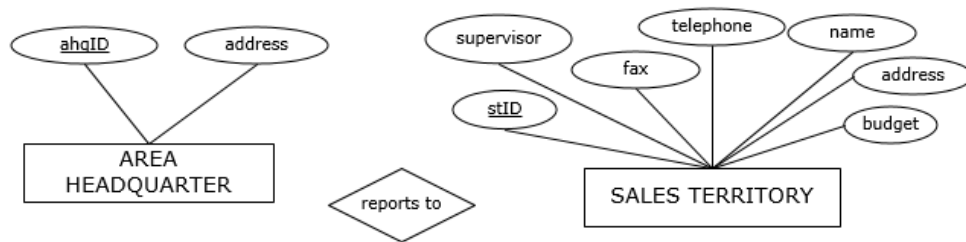
- (a) Only Row #1  
 (b) Only Row #3  
 (c) Only Row #1 and Row #2  
 (d) Only Row #2 and Row #3  
 (e) All of the rows are possible

### Memory Recall (Experiment 2)

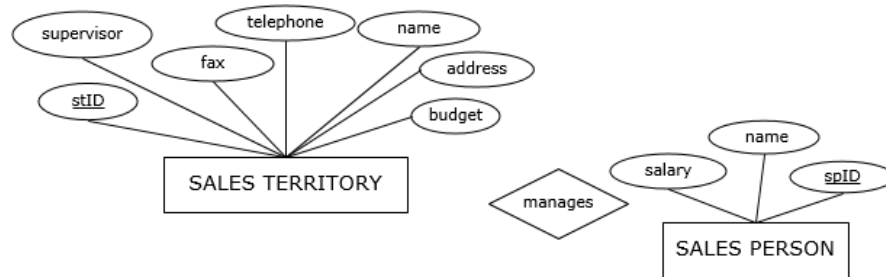
Respondents had to recall from memory the appropriate relationship cardinality in their schema and correctly select a true/false response. The set diagram was advantageous for the respondents because they could see examples which made the semantic more concrete before the diagram was removed. The increased perceptual information of the instances encourages the individual to more thoroughly consider the cardinality constraint rules for the entire entity type in a manner that is more natural to them (i.e., without symbols) and recall them later without the diagram. For example, the answer to question 1 (a) and (b) (mandatory cardinality) was “false” and “true,” respectively. This question required the respondents to examine the relationship between PRODUCT LINE and WAREHOUSE. For respondents who received the set diagram, they could map the concrete examples directly to the question asked. On the other hand those who received the ER diagram had to recall the less concrete symbols of the schema in their head, thus not benefiting from the concreteness effect.



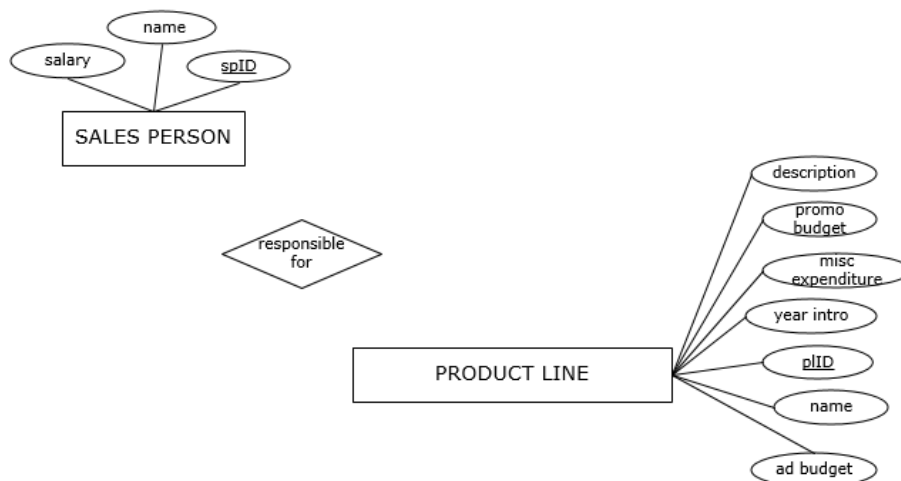
### Optional Cardinality: ERD



1. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "reports to"?
  - (a) An area headquarter can be associated with a minimum of zero sales territories. True / False
  - (b) A sales territory must be associated with a minimum of one area headquarter. True / False

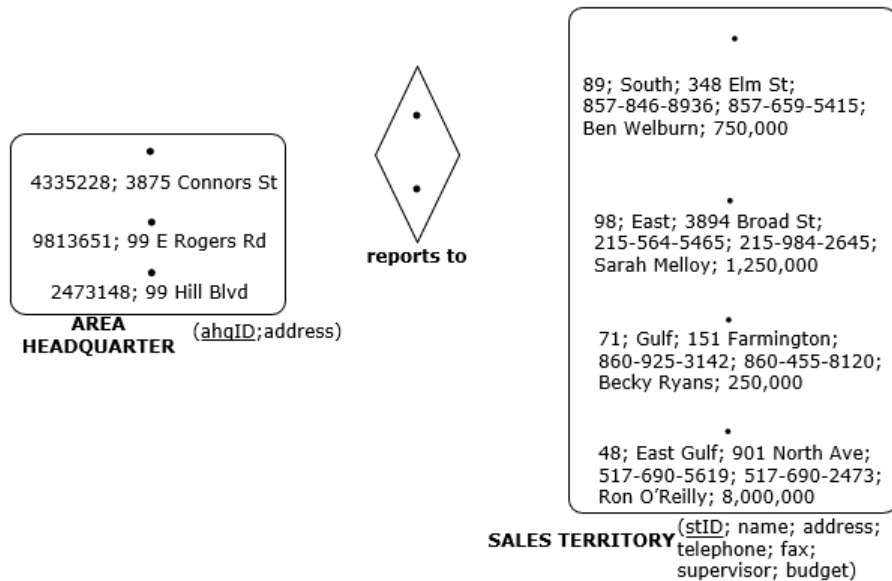


2. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "manages"?
  - (a) A sales territory can be associated with a minimum of zero sales persons. True / False
  - (b) A sales person must be associated with a minimum of one sales territory. True / False

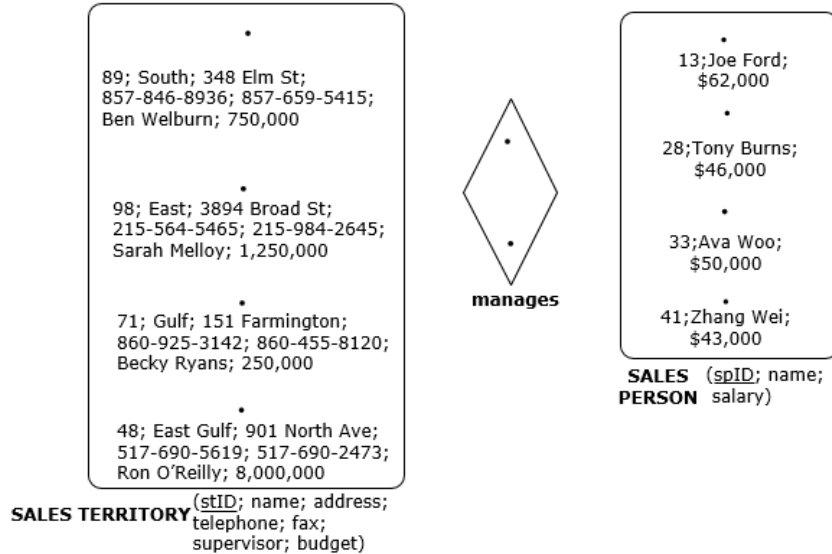


3. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "responsible for"?
  - (a) A sales person can be associated with a minimum of zero product lines. True / False
  - (b) A product line must be associated with a minimum of one sales person. True / False

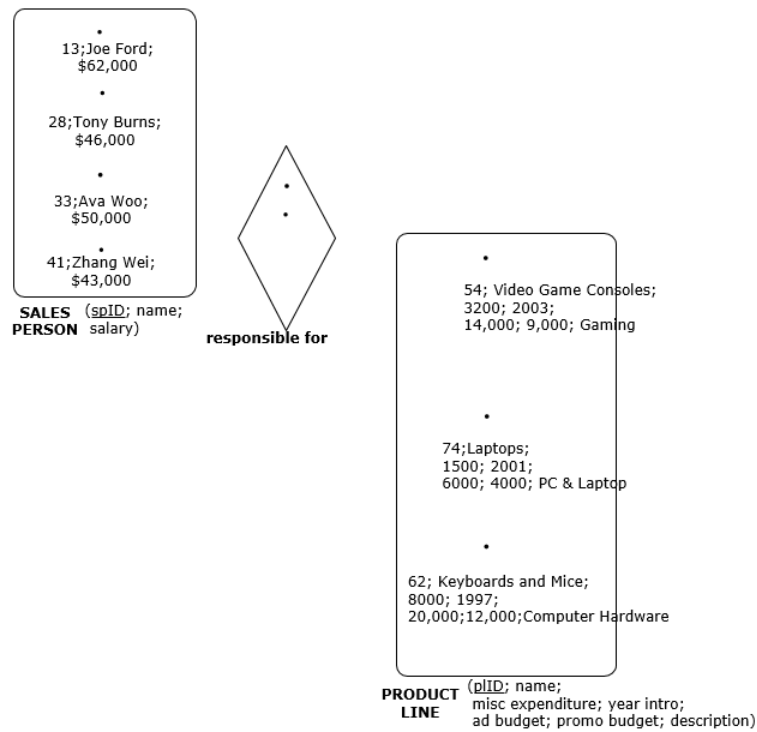
## Optional Cardinality: Set



1. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "reports to"?
  - (a) An area headquarter can be associated with a minimum of zero sales territories. True / False
  - (b) A sales territory must be associated with a minimum of one area headquarter. True / False

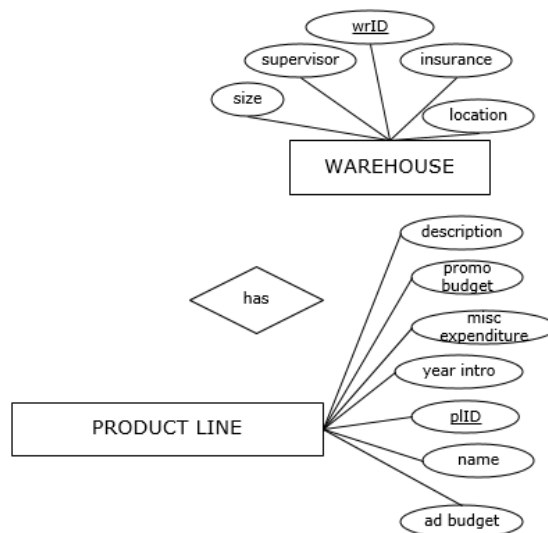


2. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "manages"?
  - (a) A sales territory can be associated with a minimum of zero sales persons. True / False
  - (b) A sales person must be associated with a minimum of one sales territory. True / False

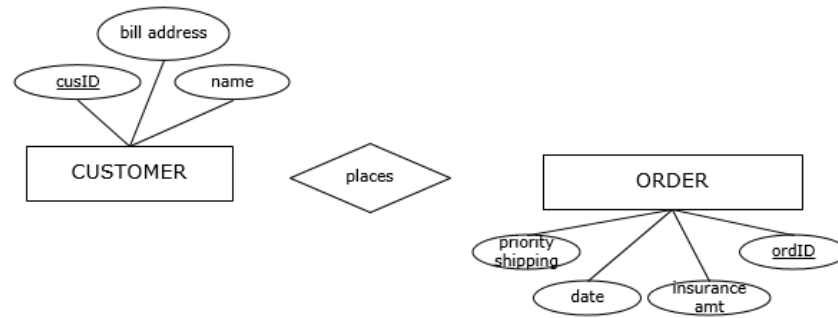


3. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, “responsible for”?
- (a) A sales person can be associated with a minimum of zero product lines. True / False
  - (b) A product line must be associated with a minimum of one sales person. True / False

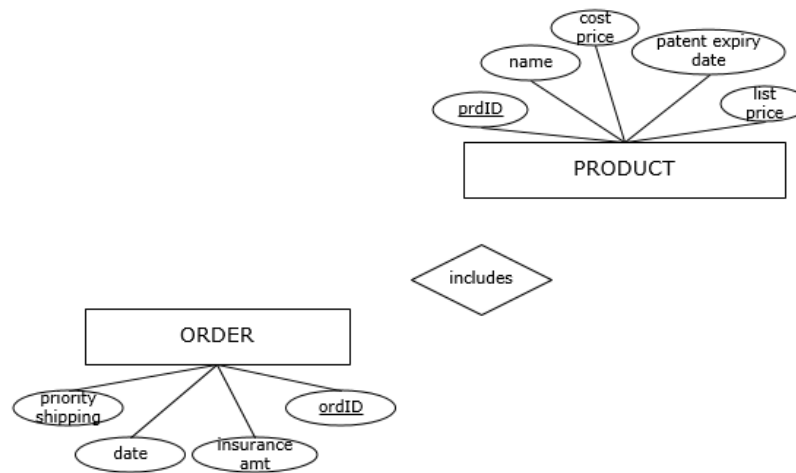
### Mandatory Cardinality: ERD



1. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, “has”?
- (a) A warehouse can be associated with a minimum of zero product lines. True / False
  - (b) A product line must be associated with a minimum of one warehouse. True / False

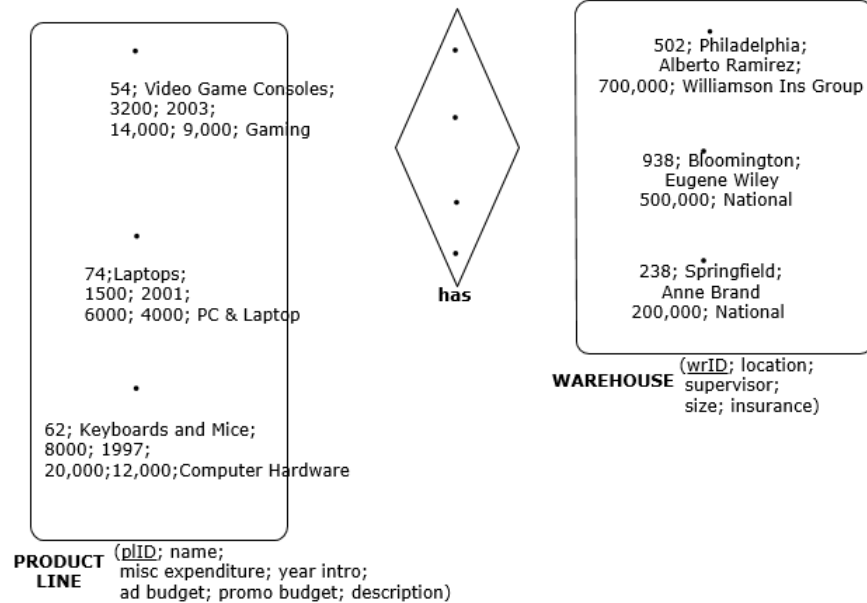


2. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, “places”?
- (a) A customer can be associated with a minimum of zero orders. True / False
  - (b) An order must be associated with a minimum of one customer. True / False

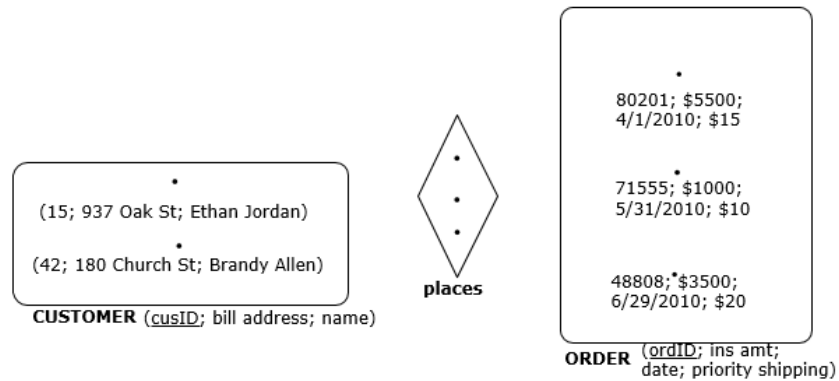


3. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, “includes”?
- (a) A product can be associated with a minimum of zero orders. True / False
  - (b) An order must be associated with a minimum of one product. True / False

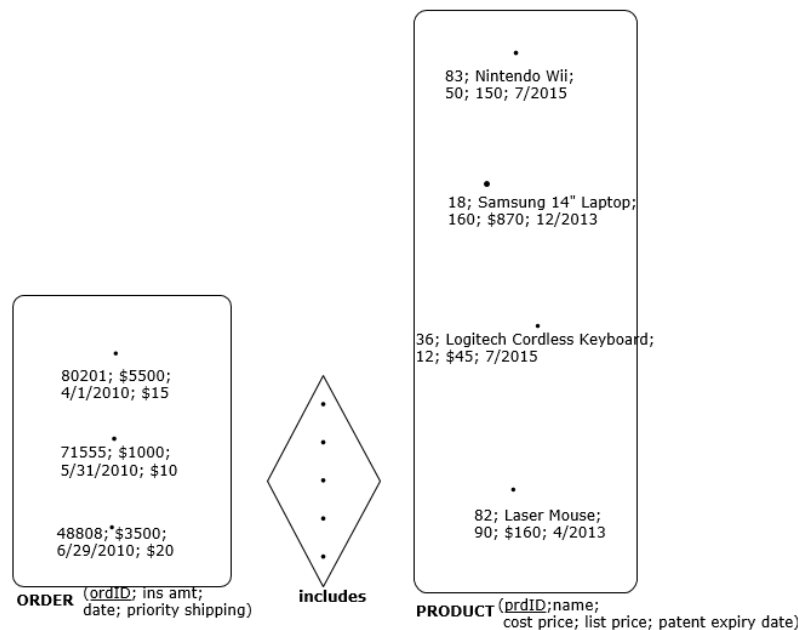
## Mandatory Cardinality: Set



1. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "has"?
  - (a) A warehouse can be associated with a minimum of zero product lines. True / False
  - (b) A product line must be associated with a minimum of one warehouse. True / False



2. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, "places"?
  - (a) A customer can be associated with a minimum of zero orders. True / False
  - (b) An order must be associated with a minimum of one customer. True / False



3. An incomplete fragment of the diagram you viewed earlier is shown above. Based on your memory of the complete diagram, which of the cardinality constraints are true for the relationship, “includes”?
  - (a) A product can be associated with a minimum of zero orders. True / False
  - (b) An order must be associated with a minimum of one product. True / False

### Schema Based Problem Solving (Experiment 2)

Respondents had to answer a question about a relationship cardinality in their schema using a scenario of a change or potential change in the business presented to them. They were instructed to choose among three options (possible, not possible, or not sure) and they were also asked to give an explanation for their choice. For example, the answer for question 1 (optional cardinality) was “possible.” This question required an individual to consider the relationship between SALES PERSON and SALES TERRITORY. Individuals who received a set diagram could employ ready-made examples to conclude that every sales person does *not* need to manage a sales territory. The set diagram directly provides ready-made examples which an individual can employ to test their understanding of the task and provide an apt explanation. Those who were given the ER Diagram had to create an example (extension of the schema), which required transformation.

#### Optional Cardinality

1. There are rumors that a sales territory which is being managed by a sales person is being consolidated. Does the diagram allow the sales person to be employed without managing a sales territory?
2. Company shareholders want data about every product line’s ad budget and it has been decided that the sales person responsible for the product line should report its ad budget yearly. Is it possible that some product lines might not have their ad budget reported?
3. Due to the years of internal company knowledge required to do the job properly, product line managers are a difficult position to staff. Does the diagram allow for a product line to not be managed by a product line manager?

#### Mandatory Cardinality

1. Customers tend to modify and cancel orders. Does the diagram allow an order to not have an associated customer?
2. The dynamics of the market require products to often be realigned to another product line. Can a product not belong to a product line?
3. A new product line is being created and will eventually have many warehouses. Can the new product line be launched without any of its warehouses selected?

# Appendix F

## Schema-Based Problem-Solving Explanation Coding Scheme

For each question, participants were asked to choose among three options (possible, not possible, or not sure) and give an explanation for their choice. Each explanation needs to be evaluated on two criteria, data identification and logic/reasoning, in determining an overall score:

1. *Data Identification*: the extent to which the respondent derived relevant information from the schema/model.
2. *Logic/Reasoning*: how well the response is constructed, logically, and describes cardinality.

After evaluating the responses on these two criteria, assess the respondent's overall response. Priority is given to logic/reasoning.

### Data Identification

Based on the aspects that should have been included in the answer, what is the extent to which the crucial pieces of data are included in the answer?

| Code           | 0   | 1   | 2  |
|----------------|---|---|--|
| Interpretation | No specific mention about the entity type, entities, relationship type, or relationship instances from the diagram mentioned. | <ul style="list-style-type: none"> <li>Only a partial mention of some (e.g., one side) of the relevant entity type, entities*, relationship type, or relationship instances* from the diagram.</li> </ul> | <ul style="list-style-type: none"> <li>Mentions both sides of the relevant entity type, entities*, relationship type, or relationship instances*.</li> </ul> |

Notes: \*Does not need to mention all of the entities or relationship instances

### Logic/Reasoning

How convincing is the logic/reasoning for obtaining the information?

| Code           | 0  | 1  | 2   |
|----------------|--|--|---|
| Interpretation | No reasonable response about cardinality rules provided. Also for individuals that stated they did not know. | <ul style="list-style-type: none"> <li>Only a partial correct response about cardinality is provided.</li> <li>The logic/reasoning is incomplete.</li> </ul> | <ul style="list-style-type: none"> <li>A correct response about the cardinality rule is provided.</li> <li>The logic/reasoning is appropriate.</li> </ul> |

### Overall

Overall judgment of quality evaluates the overall answer. This score will take into consideration both data identification and reasoning, along with the coder's best judgment. Scoring this item can therefore be regarded as holistic in nature. You have the right to overwrite/give another score if you think it is more appropriate.

| Overall Score       | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------------------|---|---|---|---|---|---|---|---|---|
| Logic/Reasoning     | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 |
| Data Identification | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |

Copyright of MIS Quarterly is the property of MIS Quarterly and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.