An Experimental Examination of Property Precedence in Conceptual Modelling

Jeffrey Parsons and Linda Cole
Faculty of Business Administration
Memorial University of Newfoundland
St. John's, NL Canada
jeffreyp@mun.ca, lmcole27@yahoo.ca

Abstract

Interest in evaluating conceptual modelling techniques has recently experienced a revival, in part due to widespread adoption of the Unified Modelling Language (UML). In addition, the use of ontology as a framework for evaluating conceptual modelling techniques has gained acceptance. In this paper, we consider implications of applying one aspect of the ontology of Mario Bunge to conceptual modelling. Specifically, conceptual modelling has traditionally failed to provide mechanisms to indicate that some properties of types or classes may be considered dependent on others. This paper presents a theoretical rationale, using Bunge's ontological notion of *precedence*, for explicitly modelling such dependence in conceptual schema diagrams. We present the design of an experimental framework to test the impact of explicitly representing precedence on the ease with which a diagram can convey domain semantics. In addition, we consider how the issue of "common sense" semantics can interfere with experimental procedures to evaluate the semantics conveyed in a diagram's structure. We offer early experimental results indicating: 1) the explicit modelling of precedence improves the ability of experimental participants to verify the existence of dependence among properties (but has no effect on the ability to verify the semantics conveyed by association cardinalities); and 2) the potential for background knowledge to interfere with the semantics conveyed by diagram structure. We conclude by discussing the need for further research on both these issues.

Keywords: property precedence, ontology, experimental evaluation of conceptual modelling techniques.

1 Introduction

According to some accounts (e.g., Oei et al. 1992), hundreds of modelling methods and techniques for systems analysis and design have been proposed by researchers and practitioners. However, few have seen widespread use in practice. Popular methods focus on data (e.g., ER diagrams) or processes (e.g., data flow diagrams), or use combinations of diagramming techniques to capture the structural and behavioural dimensions of objects (e.g., UML class diagrams and collaboration diagrams). In addition, there have been numerous attempts to extend conceptual modelling techniques to capture additional domain semantics.

In the context of these developments, a common criticism of modelling techniques is that they lack strong theoretical foundations regarding the subject matter they model (Wand and Weber 1993, 2002). Since techniques generally lack theory, it is difficult to understand why particular extensions are proposed, and if and how new techniques or extensions of existing ones are useful in supporting specific conceptual modelling objectives. This challenge has generated a steady stream of research aimed at evaluating techniques, particularly those involving diagrammatic notations that convey semantics of a real world domain.

In this paper, we propose an extension to UML class diagram notation intended to more directly express the semantics associated with the ontological concept of *property precedence* (Parsons and Wand 2002) – the notion that properties can be expressed in a hierarchy from more general to more specific. We then discuss an experiment to test the impact of including precedence semantics in class diagrams and present results of a preliminary experimental study. We also consider how 'common sense' semantics conveyed by the words in a diagram can interfere with the semantics conveyed by the structure of the diagram. We conclude by summarizing the substantive and methodological issues in this research, and discussing future research opportunities.

2 Theory

2.1 Related Prior Research

Recently, a number of studies have looked at using ontological foundations as the basis for evaluating conceptual modelling techniques. For example, Burton-

Jones and Weber (1999) studied relationships with attributes in entity relationship diagrams. They hypothesized that relationships with attributes are ontologically unclear, undermining the understanding of a conceptual model and impairing the ability to solve problems about the domain. They found partial support for their hypotheses; in an unfamiliar domain (plant nursery), ontological clarity (relationships without attributes) had a positive impact on problem solving performance.

Gemino and Wand (2000) compared mandatory and optional properties in conceptual modelling. They hypothesized that, since optional properties are inconsistent with ontological principles, problem solving scores would be significantly higher for the mandatory properties group then for the optional properties group. Their results supported the proposition that grammars using subtypes with mandatory properties lead to a higher level of understanding that those using optional properties.

Bodart et al. (2001) also compared mandatory and optional properties in conceptual modelling and found similar results. They hypothesized *first* (second) that diagrams using *optional properties* (mandatory properties) would assist users undertaking tasks that required a *surface level understanding* (deep level understanding) of the domain better than those using only mandatory properties (optional properties). The results supported the hypothesis. Bodart et al found that the optional properties group outperformed the mandatory properties group in diagram reconstruction and comprehension questions; however, the mandatory properties group outperformed the optional properties group in problem solving questions.

Burton-Jones and Weber (2003) most closely approached the issue of interest in this paper. They investigated the practice of representing properties of properties in conceptual modelling, notably properties of relationships in conceptual schema diagrams. Based on Bunge's (1977) ontological claim that "properties do not have properties," they hypothesized that modelling relationships with attributes would decrease both understanding of conceptual models and confidence in ability to interpret model semantics. They found support for the hypotheses related to comprehension, but found no effects on confidence.

Burton-Jones and Weber (2003), however, did not draw on the notion of precedence to support their predictions about "properties of properties." Next, we outline the concept of property precedence as a basis for making competing predictions about the value of modelling properties of properties.

2.2 Ontological Foundation

We start from the fundamental premise that conceptual models represent human conceptions of aspects of the real world. We also adopt the view that ontology, the branch of philosophy dealing with the nature of reality, can provide insights into understanding the real world phenomena that need to be represented in conceptual models. In particular, we follow Wand and Weber (1990, 1993, 2002, Weber 1997) in taking the position that the ontology of Mario Bunge (1977, 1979) provides a highly formalized and useful basis for understanding conceptual modelling. In this section, we present the ontological basis for articulating hierarchical relations among properties possessed by things. The definitions that follow are adapted from Parsons and Wand (2002).

Postulate*:[1] The world is made of *things* that possess *properties*.

This postulate can be viewed as the underlying ontological justification for including in a conceptual modelling technique a construct to represent properties; namely, conceptual models represent the properties of things. The property (or attribute construct) is widely used in conceptual modelling.

An important distinction can be made between two kinds of properties. *Intrinsic properties* can be understood in terms of a single thing in isolation from other things in the world. For example, the 'weight' of some object is generally independent of other objects in the world. *Mutual properties*, however, are fundamentally properties that have meaning only in the context of two or more things. For example, 'owns' is a mutual property that is meaningful only in the context of at least two things (the owner and the owned).

In a traditional conceptual modelling approach such as the Entity Relationship model, intrinsic properties can be represented as attributes of entity types, while mutual properties can be represented as relationships between entity types (Wand, Storey & Weber 1999).

Principle*: There are no things without properties and *properties are always attached to things* (Bunge 1977, pp. 36, 58, 62).

Since, in Entity-Relationship based conceptual models, properties (attributes and relationships) provide the basis for determining the similarity of things in establishing entity types or classes – when organizing information there is a special interest in identifying which things possess which properties.

Definition*: The *scope* of a property, P, is the set of things possessing the property, denoted Scope(P).

2.3 Property Precedence

Bunge recognizes that properties that appear to be different at one level can sometimes be generalized and regarded as the same at a more abstract level. The connection between specific and generic properties can be formalized via the notion of *property precedence*.

Definition*: Let P_1 and P_2 designate two properties. P_1 will be said to *precede* P_2 iff for every thing x possessing P_2, x also possesses P_1.

[1] The * indicates postulates and definitions adapted from Bunge's model by Parsons and Wand (2002).

Note that P_1 precedes P_2 if and only if the $Scope(P_2) \subseteq Scope(P_1)$.

For example, the property 'make sound' precedes the properties 'talk' and 'bark' since every thing that can talk and every thing that can bark can make sound. The set of things that bark is a subset of the set of things that make sound.

Definition*: Let ***P*** denote a set of properties. A *preceding property* of a property P in ***P*** is a property that must be possessed by any instance possessing P.

Note that, if property Q is a preceding property of property P, this can be expressed by the equivalent statements, P *is preceded by* Q, or Q *precedes* P.

The following are examples of precedence applied to intrinsic and mutual properties, respectively:

1. To be 'born May 1, 1970' is preceded by to 'have date of birth.'
2. To 'have a contract termination date between a customer and supplier' is preceded by to 'have a contract between a supplier and a customer.'

As noted above, if Q precedes P then $Scope(Q) \supseteq Scope(P)$. Thus, we can state:

Definition*: Let $\wp(\boldsymbol{P})$ denote the power set (set of all subsets) of ***P***. The *preceding properties* of P in ***P*** are defined by the function Preceding: $\boldsymbol{P} \rightarrow \wp(\boldsymbol{P})$, such that $Preceding(P)=\{Q \in \boldsymbol{P} \mid Scope(Q) \supseteq Scope(P)\}$.

Definition*: The *preceded properties* of a property P are all properties for which P is a preceding property. That is, the function Preceded: $\boldsymbol{P} \rightarrow \wp(\boldsymbol{P})$ such that $Preceded(P)=\{Q \in \boldsymbol{P} \mid Scope(P) \supseteq Scope(Q)\}$.

Typically, a class (or type) in a conceptual model is represented in terms of a set of attributes possessed by all instances of the class. Precedence has special importance in the context of the relationship between properties and classes (or entity types). Often, classes are described by some *generic* properties, while the instances possess *specific* attributes implying the generic ones (typically not shown in a conceptual schema constructed using a modelling grammar such as the ER model). We say that the specific property *manifests* the generic one. Different instances might possess different properties preceded by the same generic property.

According to Parsons and Wand (2002), there are two main types of manifestation precedence. *Value precedence* occurs when a given generic property is manifested by a specific value. For example, having a specific weight is preceded by the property 'has weight.' Each instance having a specific weight value can be a member of a class that includes in its definition 'has weight.' In this case, the value has no meaning unless "attached" to the generic property (consider the example of weight). *Specialization precedence* occurs when the generic property can be specialized in different ways where the individual cases do have meaning independent of the more generic property. For example 'moves on land' may be manifested as 'crawls,' 'walks,' 'hops,' or 'runs.' Animals propelling in one of these ways can be members of a class that includes in its definition the property 'moves on land.'

In both cases, there is a *dependence* between the preceded and preceding properties, in which possessing the preceded property implies possessing the preceding property.

We believe that manifestation can play an important role in conceptual modelling. Common modelling approaches, such as the ER model or UML class diagrams, do not provide constructs to capture precedence. Value precedence is generally not an issue, since these models deal with class-level issues (i.e., they do not model instances). Specialization precedence, which can arise in the context of attributes or associations/relationships, is an issue, but is typically not dealt with in conceptual modelling, except to the limited extent that class hierarchies indicate that instances of a subclass are also instances of a superclass.

This paper seeks to examine how precedence might affect the ability to verify the semantics conveyed in a conceptual schema. Thus, we propose:

***Proposition*:** Extensions to conceptual modelling techniques that allow specialization precedence to be directly expressed will communicate domain semantics more effectively than semantically equivalent diagrams that do not express specialization precedence.

Note that this proposition directly contradicts the above-mentioned proposition by Burton-Jones and Weber (2003) that "properties should not have properties." We believe that their analysis, while based on Bunge's ontology, fails to account for the possible interpretation of a preceded (manifested) property as a "property of" a more general preceding one. Thus, the present work contributes to the ongoing body of work seeking to improve our understanding of conceptual modelling phenomena.

2.4 Capturing Precedence in Class Diagrams

We propose a minor extension to UML class diagram notation to allow the direct expression of precedence. Here, we are less concerned with the details of the extension than in using the extension as a treatment in an experimental comparison of whether the inclusion of precedence semantics affects the way in which class diagrams convey domain semantics. Such an extension could be applied to any conceptual modelling language that has constructs for modelling properties. Hence, we do not propose this extension as a formal extension to UML notation that could be considered for adoption by the OMG. However, we believe that, in general, the results of such evaluations should be considered by those who control the evolution of the UML.

In addition, we restrict our attention to specialization precedence as applied to *associations*. Similar extensions could be applied to deal with specialization or value precedence as applied to attributes.

In class diagrams, associations are depicted using labelled lines between classes, indicating links between instances of the classes connected by the association. If additional information about the association needs to be captured, an association class (which may have attributes that express more semantics of the association) may be linked to the association. Figure 1 illustrates the use of associations and association classes with attributes. Our notation differs somewhat from standard UML in that we attach labels to the various lines linking the object and association classes to better describe the links. For example, in Figure 1 the link "Takes" describes a generic association between Students and Courses, while "Enrols in" describes more specifically the nature of the "Takes" association between Students and Courses via the association class "Section."

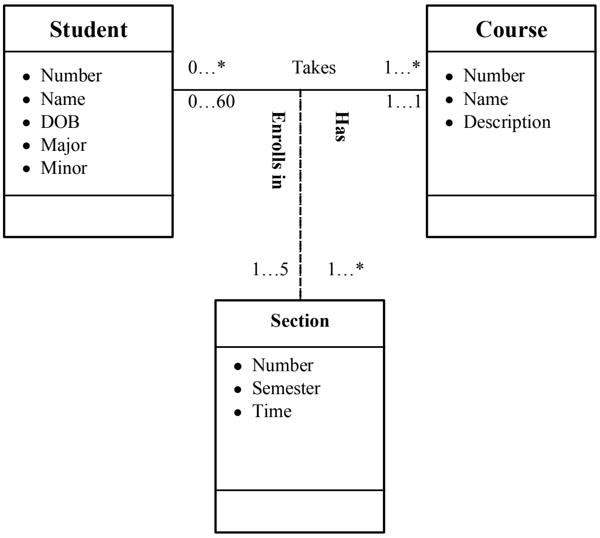

Figure 1: Associations and Association Classes

To illustrate precedence, we extend the example in Figure 1 to consider the case where a course has lab requirements, and a lab is associated with a section. This reflects specialization precedence in the sense described earlier, since the enrolment in a lab (i.e., the association of a student with a course through a lab) implies a more "generic" association between a student and a course though a section. Traditionally, one might model this through a series of associations, or by elevating "Section" from an association class to an object Class. In class diagrams, there are no theoretically-grounded prescriptions to distinguish situations calling for modelling phenomena using object classes from those calling for modelling phenomena using association classes. We propose, instead, to express the semantics of precedence by allowing secondary (this can be expanded to as many level as necessary) association classes to be linked with other (primary) association classes. Figure 2 illustrates this.

In Figure 2, the attachment of Lab to Section makes explicit the view that a lab is a mutual property between a Student and a Course that is preceded by another more general mutual property, Section.

Our research, then, focuses on whether the direct expression of precedence semantics (e.g., as shown in Figure 2) improves the ease with which the semantics of precedence can be communicated in a conceptual model, relative to an informationally equivalent representation that does not directly express precedence semantics. The next section describes the design of an experiment to answer aspects of this question.

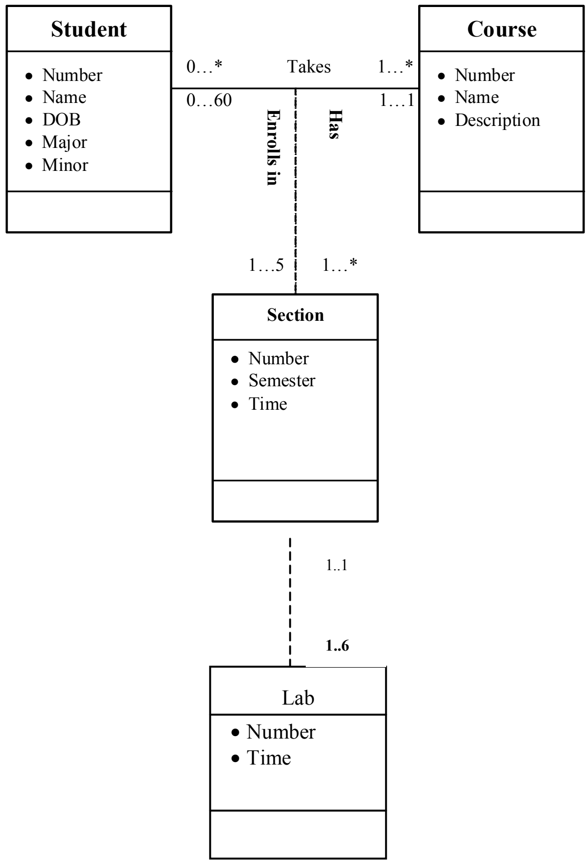

Figure 2: Expressing Specialization Precedence via Association Classes

3 Research Method

3.1 Experimental Design Issues

Our proposition that diagrams that enable precedence to be expressed directly will communicate domain semantics more effectively than those that do not raises a number of interesting issues with respect to experimental design. First, there may be many ways of expressing domain semantics "directly." It is not the purpose of this research to compare alternative diagram formats that each directly express precedence, since the underlying ontological framework does not speak to that issue. Instead, focusing on precedence relations among mutual properties, we use the approach outlined in Section 2.3 as our primary treatment (direct representation of precedence involving mutual properties). We use as the indirect representation a standard UML diagram in which mutual properties are represented as distinct associations.

Second, when comparing representation formats, alternative conceptual schema diagrams generated to manipulate the **independent variable** must be *informationally equivalent* (Gemino and Wand 2003). That is, it should be possible to answer questions correctly with any of the representational forms used as treatments in an experimental study. Otherwise, the questions chosen to measure dependent variables predetermine the results. To illustrate, if one form provides enough information to answer selected questions correctly, while a second form does not, it would be neither surprising nor interesting to find that participants receiving the first form outperform those receiving the second form on those questions.[2] This does not mean there needs to be strict information equivalence between diagrams; however, with respect to questions asked related to the dependent variables, the representations must provide all information needed to answer the questions correctly. In our study, we handle the information equivalence issue by ensuring in our choice of questions that the alternative representation formats allow those questions to be answered correctly.

Third, **dependent variables** should *measure performance only with respect to semantics contained in the conceptual schema diagrams.* Therefore, questions should be limited to those that test understanding of explicit semantics expressed in a script. Other kinds of questions, such as the problem solving questions used by Gemino (1999) and Bodart et al. (2001), necessitate prior background knowledge to answer meaningfully or correctly, and therefore do not deal with communicating/validating domain semantics.[3] We deal with this in our study by developing only questions that require nothing more than verifying the semantics expressed by the structure of a diagram.

Fourth, when selecting **participants** for studies testing the ease with which alternative forms of representation convey domain semantics, *subject matter experts (SMEs) should not be used*. This may seem a strange recommendation, since a central role of conceptual models is to facilitate communication between analysts and users (SMEs). However, in this study we are interested in the degree to which (and ease with which) information can be extracted from a script. Therefore, it is critical that participants be able to answer questions by using only that script, rather than by using some background knowledge (e.g., experience with a domain similar to that represented in a script). Using SMEs in tests of the capacity of scripts to convey semantics can clearly confound experimental results. To illustrate, if we construct a script related to the enrolment of students in courses at a university and conduct an experiment with student subjects (reasonably viewed as SMEs with respect to enrolment in courses at a university), there is no way of distinguishing whether answers are based on information provided in the scripts, or simply on the participants' experience at their own university. This can weaken or strengthen the effects of the treatment, depending on whether or not the participants' knowledge is consistent with the semantics in the diagram.

Clearly, tests involving SMEs are needed in order to fully understand issues in using conceptual modelling techniques to communicate and verify requirements. However, we argue that such tests should follow the development of a reasonably comprehensive understanding of how different representations facilitate or inhibit the expression of semantics. We further contend that such understanding does not yet exist, and that the design of some prior studies may have inhibited the development of cumulative knowledge in this area.

To deal with this issue, in this study we use alternate forms of the experimental material. In one form, we use English words suitable to a domain with which most or all participants will have at least some familiarity. In the second version, we substitute Greek letters so that participants are able to answer questions only on the semantics conveyed by the diagram structure.

Therefore, a second objective of this study is to determine whether background knowledge can interfere with interpretations of the semantics expressed by a diagram's constructs. Some prior research (Siau et al. 1997) has suggested that this is not the case, but that work focused very narrowly on a contrived situation involving conflicts between semantics expressed by cardinality constraints and common sense semantics implied by the labels attached to entity types and relationships in an Entity Relationship diagram.

Finally, in **experimental procedures** for studies testing the representation capacity of alternative forms of representation, scripts should be *available to participants for review as they answer questions used to measure the dependent variables*. Since we are interested in testing the ability of the diagrams to convey semantics and since scripts would be available in a real-world communication/validation situation, there is no reason to remove diagrams in experiments. Some have advocated removing diagrams on the basis that an objective of studies may be to test 'learning' (Bodart et al 2001, Gemino and Wand 2003). However, this objective is difficult to reconcile with the role of conceptual models in facilitating communication with users/clients who presumably know the domain, and would not learn from reviewing diagrams. Testing learning may lead to important future uses for conceptual models, but we believe our understanding of how methods convey domain semantics remains sketchy to the point that considerable work is needed in understanding more basic issues first.

---

[2] Note that the issue of how easy or natural it is to accommodate certain information within a representational form is a separate question. Some forms may better facilitate the representation of certain semantics than others. We contend that 'writing' tasks are more suitable for answering such research questions than are reading tasks.

[3] It may be appropriate to use problem solving to test how well someone has assimilated an unfamiliar domain from a diagram. This is an interesting question, but does not deal with the main purposes of conceptual modelling and is, therefore, beyond the scope of this paper.

3.2 Task Domain and Material

Since the objectives of this study are very focused, we decided not to test entire diagrams. Instead, we used several small diagram segments focusing on specialization precedence among mutual properties. Figures 3 and 4 illustrate alternate representations for one of the diagram segments used. Both diagrams model logical associations (flight and reservation) between physical objects (passengers and aircraft). In Figure 3, the idea that reservation is preceded by flight, and that flight is a mutual property linking person and aircraft, is not shown. Figure 4 introduces the additional notation we use to depict precedence. Flight is shown as an association class, and Reservation is shown as a secondary association. This means that the existence of a reservation between a passenger and an aircraft implies the existence of a flight linking passengers and aircraft. Note that we recognize alternative equivalent representations are possible for this scenario.

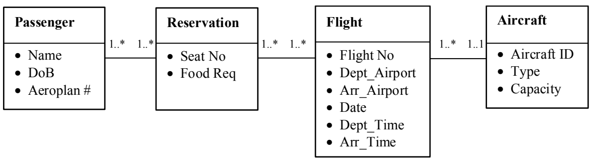

Figure 3: No Representation of Precedence

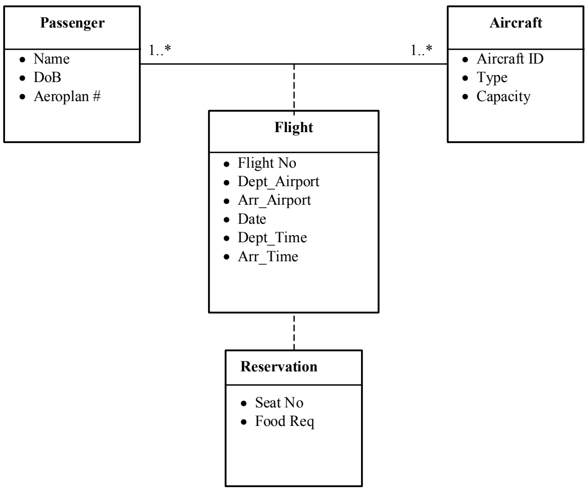

Figure 4: Direct Representation of Precedence

In addition, to test whether background knowledge interferes with the communication of diagram semantics, we constructed alternative representations for which the class and association labels were "semantically void." Figures 5 and 6 contain segments corresponding to those in Figures 3 and 4, respectively.

As can be seen from Figures 5 and 6, answers to any questions about the diagrams can be based only on the semantics conveyed via the structure of the diagram. In contrast, in Figures 3 and 4, answers might be confounded by background knowledge about the general subject domain – airline reservations.

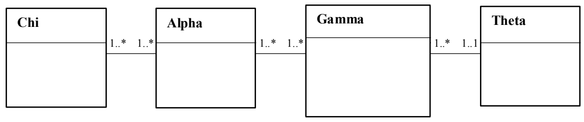

Figure 5: Semantically Void Equivalent of Figure 3

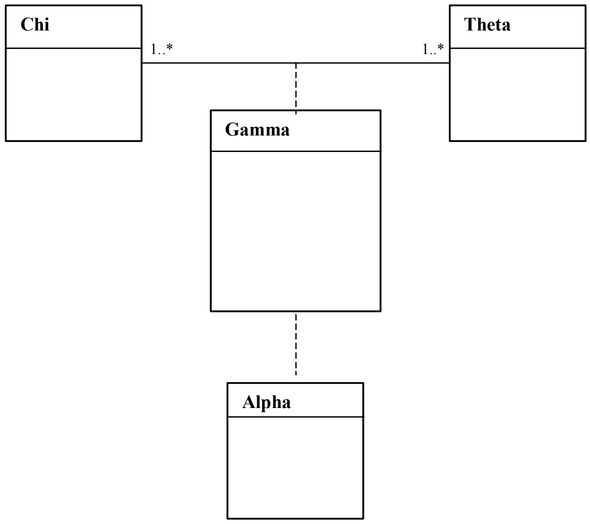

Figure 6: Semantically Void Equivalent of Figure 4

The dependent variables in this study consist of the number of correct answers to questions about the semantics conveyed in a diagram segment. Table 1 contains the questions asked regarding the diagrams in Figures 3/4 and 5/6.

Semantic Segment (Figures 3/4)	Semantically Void Segment (Figures 5/6)
Can a Passenger have a Reservation with an Aircraft that does not involve a specific Flight?	Can a Chi be associated with an Alpha that does not involve a specific Gamma?
Does the diagram specify an upper limit on the number of Reservations associated with a specific Flight?	Does the diagram specify an upper limit on the number of Alphas associated with a specific Gamma?

Table 1: Sample Questions

The questions for both examples involve two aspects of semantics. In the example shown here, the first question deals with whether a particular association can exist without another association. More generally, the semantics here involve the existence of precedence. In Figure 4, this is shown directly, while in Figure 3 it has to be inferred through a chain of associations linking object classes. Although a correct answer can be provided on the basis of the information in either diagram, the depiction in Figure 4 is more direct; therefore, we expect it to be more easily verifiable.

The second question involves the semantics carried in the cardinality constraints.[4] Although this does not deal with precedence per se, and the information appears relatively the same in both diagram segments, we include an example of this type of question because of the importance of cardinality constraints in general in conceptual modelling.

The experimental materials consisted of four diagram segments (two subject domains, direct representation of precedence vs. indirect representation of precedence), each with a series of questions. In addition, we also varied the use of segments that carried semantics in the words used in class and association labels (which we call Semantic segments) with the use of segments that used Greek letters as labels and were therefore semantically void (which we call Void segments). Two versions of the material were developed. In version A, the ordering of segments was: Void-Indirect; Void-Direct; Semantic-Direct; Semantic-Indirect. In version B, the ordering of segments was: Void-Direct; Void-Indirect; Semantic-Indirect; Semantic Direct. In both versions, the Void cases were presented first to eliminate the potential influence of semantics conveyed by the use of words on subsequent interpretation of semantically void segments.

Using this counterbalanced approach, we were able to double our effective sample size for a given number of participants.

3.3 Participants and Procedure

The participants in this study were 29 computer science students in an undergraduate database course. In this course, all participants had completed three weeks of material on conceptual modelling using UML class diagrams. However, the extended notation used in this experiment to represent precedence had not been covered. Instead, participants were simply told to make their best interpretation of the semantics conveyed by having relationship classes attached to relationship classes in order not to focus attention on the purpose of the study.

Each participant was randomly assigned to version A or version B (as described above) of the material. Since each participant saw four conceptual schema diagram segments, this resulted in an overall N of 116.

The material was distributed at the beginning of a review class at the end of the semester. Participants were told that the purpose was to test how well they understood the semantics expressed by the modelling constructs in class diagrams. No course credit was given for the exercise. Nevertheless, participants were motivated to take the task seriously in view of an upcoming final exam in the course. Booklets containing instructions, the diagram segments, and questions for each segment, were distributed. Participants were given as much time as needed to complete the exercise and the instructor collected each booklet as it was completed. After all completed booklets had been collected, the instructor reviewed the questions and gave correct answers with explanations.

3.4 Operational Hypotheses

To test the general proposition outlined above, the following specific hypotheses were developed:

H1a: Scores on questions about *existence of precedence* will be higher when diagrams contain Direct representation of precedence than when they contain Indirect representation of precedence.

H1b: Scores on questions about *cardinality semantics* will be higher when diagrams contain Direct representation of precedence than when they contain Indirect representation of precedence.

These hypotheses are driven by the importance of precedence in Bunge's ontology. In essence, we expect that diagrams that include this concept will communicate semantics more effectively than those that do not.

To test whether the informal semantics conveyed by the use words to label classes and associations can interfere with the semantics conveyed by the constructs used in a diagram, the following hypotheses (parallelling the structure of H1) were developed:

H2a: Scores on questions about *existence of precedence* will be different when diagrams are semantically Void than when they contain words that can be given a Semantic interpretation based on past knowledge.

H2b: Scores on questions about *cardinality semantics* will be different when diagrams are semantically Void than when they contain words that can be given a Semantic interpretation based on past knowledge.

This hypothesis is driven by the need to ensure that we test only the domain semantics conveyed by a diagram's structure, rather than the potentially confounding semantics that might arise from participants' interpretations, based on prior knowledge, of words using in a conceptual schema diagram. Note that H2a and H2b are non-directional. We propose only that we will find a difference, and not the direction of that difference. This could be refined to predict that scores will be higher on semantically void diagrams when the diagram structure conflicts with "common sense" semantics associated with words in a diagram, but may not be different when the diagram structure is consistent with "common sense" semantics that may be attached to words in a diagram. However in view of the results of Siau et al (1997), we do not test these refined hypotheses here.

4 Preliminary Results

As described above, there were two types of questions for each segment: a question about the existence of precedence and one or more questions about cardinalities. Each response was coded as '1' for a correct answer and '0' for an incorrect answer. The 'Restaurant' segments contained three questions about different aspects of cardinality; the scores on these were combined to give

[4] In the second example used in the study, we used three questions dealing with different aspects of cardinality.

one dependent variable. The second column of both Tables 2 and 3 describe this in more detail.

To examine H1a and H1b, we used t-tests to compare the scores on each question type for each diagram segment between the direct and no representation of property precedence. In addition, we separated scores on the semantic versus semantically void versions of the questionnaire to clarify the presentation. Table 2 contains the results of this analysis.

Diagram Segment	Question Type (Note: these are not question wordings)	Semantics	Mean		SD		Sig.
			Direct	None	Direct	None	
Airline	Existence of precedence [one question]	Semantic	1.00	0.07	0	0.067	< .001
	Cardinalities related to preceding mutual properties [one question]		0.86	1.00	0.363	0.000	n.s.
Restaurant	Existence of precedence [one question]		1.00	0.57	0	0.514	.004
	Cardinalities related to preceding mutual properties [three questions]		1.33	2.71	0.488	0.468	< .001
Airline	Existence of precedence [one question]	Void	0.80	0.79	0.414	0.426	n.s.
	Cardinalities related to preceding mutual properties [one question]		0.73	0.79	0.458	0.426	n.s.
Restaurant	Existence of precedence [one question]		1.00	0.33	0	0.126	< .001
	Cardinalities related to preceding mutual properties [three questions]		1.07	2.33	0.730	0.724	< .001

Table 2: Results for Hypothesis 1 (Direct vs. No Representation of Precedence)

Diagram Segment	Question Type (Note: these are not question wordings)	Precedence	Mean		SD		Sig.
			Semantic	Void	Semantic	Void	
Airline	Existence of precedence [one question]	Direct	1.00	0.80	0	0.414	.04
	Cardinalities related to preceding mutual properties [one question]		0.86	0.73	0.363	0.458	n.s.
Restaurant	Existence of precedence [one question]		1.00	1.00	0	0	n.s.
	Cardinalities related to preceding mutual properties [three questions]		1.33	1.07	0.488	0.730	n.s.
Airline	Existence of precedence [one question]	Indirect	0.07	0.79	0.258	0.426	< .001
	Cardinalities related to preceding mutual properties [one question]		1.00	0.79	0	0.426	.04
Restaurant	Existence of precedence [one question]		0.57	0.33	0.514	0.488	n.s.
	Cardinalities related to preceding mutual properties [three questions]		2.71	2.33	0.469	0.724	.05

Table 3: Results for Hypothesis 2 (Semantic vs. Void Diagram Segments)

Table 2 shows an interesting pattern of results. For Question Type 1 (Existence of precedence), the results support H1a in three of four cases (Airline: Void, and Restaurant: Semantic and Void). Questions of this type asked about what amounted to the existence of precedences involving mutual properties (associations).

However, for Question Type 2 (Cardinalities), the results are either not significant (Airline: Semantic and Void), or are significant in the opposite direction to H1b (Restaurant: Semantic and Void). This indicates that the direct representation of precedence does not seem to positively influence the interpretation of cardinalities. Section 5 discusses possible explanations for this finding.

To examine Hypothesis 2, we used t-tests to compare the scores on each question type for each diagram segment between the semantic and void diagram segments. In addition, we separated scores on the direct versus no representation of precedence cases. Table 3 show the results.

Table 3 contains a mixed pattern of results that offers some support for H2. There were some differences between Semantic and semantically Void diagram segments. Specifically, when precedence was represented directly, participants shown Semantic segments outperformed those shown Void segments.

When precedence was not represented, participants shown Semantic segments outperformed those shown void segments (both Airline and Restaurant examples) on questions related to cardinalities (providing some support for H2b). Meanwhile, participants shown void segments outperformed those shown semantic segments on the question related to the existence of precedence in the Airline example (there was no difference in the Restaurant example), as expected from H2a. While these patterns are not conclusive, they do offer some evidence that the use of semantically-laden terms has the potential to interfere with the semantics conveyed by diagram structure. We discuss this further in Section 5.

5 Discussion

This preliminary study does offer some evidence that the direct representation of precedence in conceptual modelling may affect the ability to verify information in a conceptual schema diagram. However, in our study, this applied only to the existence of property precedence, and not to the ability to verify information about cardinalities of associations. This suggests that cardinality interpretation is not affected by the representation (or not) of precedence, while the depiction of precedence does affect the ability to determine that some associations are related to (preceded by) other associations.

The study also provides some evidence that the informal semantics that can be inferred from words in a diagram may interfere with the semantics conveyed by a diagram's structure, since we found differences in scores among participants receiving semantic versus void diagram segments. Since we did not attempt to include cases where the diagram semantics conflicted with the "common sense" semantics of words, we cannot comment on whether such conflicts confound interpretations. However, it appears the evidence presented here may not be consistent with that presented in Siau et al (1997), which concluded that readers of diagrams focused on the semantics conveyed by cardinalities when these conflicted with the informal semantics suggested by the words used in a diagram.

6 Conclusions and Ongoing Research

To our knowledge, this is the first study to empirically examine the impact of representing precedence directly in conceptual modelling. The empirical evidence suggests two things. First, precedence matters. Whether precedence is represented or not can affect the ability to verify pertinent semantics of the problem domain. Second, the results point to a need for care in designing studies to evaluate conceptual modelling techniques. In particular, when attempting to study the effects of including additional modelling constructs, researchers need to be aware of the potential confounds that may arise when background knowledge primed by the words used in diagrams may interfere with the interpretation of the semantics conveyed by diagram structure. We argue that, for the purpose of evaluating additional semantic modelling constructs, semantically void materials should be used.

We believe this line of research can help improve our understanding of the value of constructs in conceptual modelling, and can eventually lead to a stronger foundation for practical modelling using techniques such as the UML. Currently, we are designing a more detailed study to collect additional evidence relating to the findings reported here.

7 References

Bodart, F., Patel, A., Sim, M. and R. Weber (2001): Should Optional Properties Be Used In Conceptual Modelling? A Theory And Three Empirical Tests. *Information Systems Research,* 12 (4): 2001, 384-405.

Burton-Jones, A. and Weber, R. (1999). Understanding Relationships with Attributes in Entity Relationship Diagrams. In P. De and J. DeGross (Eds.). *Proceedings of the Twentieth International Conference on Information Systems*, Charlotte, NC, 214-228.

Bunge, M. (1977): *Treatise on Basic Philosophy (Volume 3), Ontology I: The Furniture of the World*, Boston, Reidel.

Bunge, M. (1979): *Treatise on Basic Philosophy (Volume 4), Ontology II: A World of Systems*, Boston: Reidel.

Burton-Jones, A. and Weber, R. (2003): Properties do not have properties: Investigating a questionable conceptual modeling practice. In D. Batra, J. Parsons, and V. Ramesh (Eds.), *Proceedings of the Second Annual Symposium on Research in Systems Analysis and Design*, Miami, 14 pp.

Gemino, A. (1999): Empirical Methods for Comparing Systems Analysis Modeling Techniques. Ph.D. Thesis, University of British Columbia, Vancouver, CAN, 1999.

Gemino, A. and Wand, Y.: Comparing Mandatory and Optional Properties in Conceptual Data Modeling. In P. Bowen and V. Mookerjee (Eds.), *Proceedings of the Tenth Annual Workshop on Information Technologies and Systems*, Brisbane, 97-102.

Gemino, A. and Wand, Y. (2003): Foundations for Empirical Comparisons of Conceptual Modeling Techniques. In D. Batra, J. Parsons, and V. Ramesh (Eds.), *Proceedings of the Second Annual Symposium on Research in Systems Analysis and Design*, 29 pp.

Parsons, J. and Wand, Y. (2002): Property-based Semantic Reconciliation of Heterogeneous Information Sources. In S. Spaccapietra, S. March and Y. Kambayashi (Eds.), *Lecture Notes in Computer Science 2503 (ER-2002)*, Springer Verlag, 351-364.

Oei, J., van Hemmen, L., Falkenberg, E. and Brinkkemper, S. (1992): The MetaModel Hierarchy: A Framework for Information Systems Concepts and Techniques. Technical Report, Department of Informatics, Katholieke Universiteit, Nijmegen, No. 92-17, 1992, 1-30.

Siau, K., Wand, Y., and Benbasat, I. (1997): The relative importance of structural constraints and surface semantics in information modeling, *Information Systems*, 22(2-3): 155-170.

Wand, Y., Storey, V. and Weber, R. (1999): An Ontological Analysis of the relationship Construct in Conceptual Modelling. *ACM Transactions on Database Systems*, 24(4): 494-528

Wand, Y. and Weber, R. (1990): An Ontological Model of an Information System. *IEEE Transactions on Software Engineering*, 16(11): 1282-1292.

Wand, Y. and Weber, R. (1993): On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. *Journal of Information Systems*, 3: 217-237.

Wand, Y. and R. Weber, "Information Systems and Conceptual Modeling: A Research Agenda," *Information Systems Research* , 13(4): 363-376.

Weber, R. (1997): *Ontological Foundations of Information Systems*, Melbourne, Coopers & Lybrand.