Fall 8-4-2009

THE EFFECT OF DOMAIN FAMILIARITY ON MODELLING ROLES: AN EMPIRICAL STUDY

Palash Bera
Texas A&M International University, USA, palash.bera@tamiu.edu

Andrew Burton-Jones
University of British Columbia, Canada, andrew.burton-jones@sauder.ubc.ca

Yair Wand
University of British Columbia, Canada, yair.wand@ubc.ca

THE EFFECT OF DOMAIN FAMILIARITY ON MODELLING ROLES: AN EMPIRICAL STUDY

Palash Bera
Assistant Professor of Management Information Systems
Texas A&M International University
Laredo, Texas 78041, USA
palash.bera@tamiu.edu

Andrew Burton-Jones
Assistant Professor of Management Information Systems
University of British Columbia
Vancouver, BC, V6T1Z2, Canada
andrew.burton-jones@sauder.ubc.ca

Yair Wand
Professor of Management Information Systems
University of British Columbia
Vancouver, BC, V6T1Z2, Canada
yair.wand@sauder.ubc.ca

Abstract

Conceptual modelling (CM) involves analysts working with domain experts to create a representation of a domain called a conceptual model. We address two issues in CM research. The first deals with the semantics that conceptual models convey. We propose guidelines for how analysts can reflect the semantics of a "role" in a conceptual model using the extended entity relationship (EER) method. Roles are important in organizations, but analysts have little guidance about how to model them. The second issue is the extent to which readers' prior familiarity with the domain shown in a model affects their understanding of the model. We conducted a laboratory study to determine how domain familiarity affects users' understanding of conceptual models that represent roles. Our results indicate that conceptual models developed in accordance with our guidelines show roles more clearly but that the benefit of doing so depends on model readers' familiarity with the modeled domain. In particular, our guidelines will be most useful when users have moderate knowledge of the domain shown in the model. When users are very familiar with the domain, the guidelines do not seem to have much benefit. However, when users have very little knowledge of the domain, the guidelines do help to a certain extent.

Keywords: Conceptual modelling, domain familiarity, roles.

1 INTRODUCTION

Conceptual modelling is a task carried out during information systems (IS) development (Hoffer et al. 2007). It involves systems analysts working with domain experts to create a representation of the domain called a *conceptual model*. Systems analysts create conceptual models so that they can learn about the domain and communicate their understanding of it to others in the system development process, such as users, designers, and other analysts (Wand and Weber 1993). Conceptual modelling is critical because it is undertaken in the early stages of system development and it is well known that errors found in these early stages are much simpler to correct than those found in later stages.

Motivated by the importance of conceptual modelling in practice, we aim to extend two recent streams of conceptual modelling research. The first stream focuses on the *semantics* of conceptual models – the meaning that they convey – and deals with guidelines to help analysts create conceptual models that convey semantics more accurately and more clearly (Evermann and Wand 2006; Hadar and Soffer 2006). Research in this stream is driven by a belief that most conceptual modelling techniques in use (such as the entity-relationship model and business process modelling notation) include very little guidance for how they should be used to reflect organizational semantics accurately and clearly. Specifically, we propose guidelines for how analysts can reflect the concept of a "role" in a conceptual model. Roles are important in organizations, but analysts have little guidance about how to model them effectively with existing conceptual modelling techniques (Zhu et al. 2006).

The second stream of research that we aim to extend focuses on the *pragmatics* of conceptual models – the contexts in which such models are used and the effects of context on how people use them. Research in this stream has shown that the semantics that individuals interpret from conceptual models is affected by the prior knowledge of the individuals reading these models. In particular, readers' prior knowledge of both the modelling technique used and the domain shown in the model can have a significant impact (Khatri et al. 2006). This is an important issue in our study because it suggests that guidelines for creating conceptual models (such as the ones we propose for modelling roles) may be more useful for some intended readers than for others. Only one study has shown this empirically to date (Burton-Jones and Weber, (1999), although it did not study the representation of roles (it instead studied the representation of relationships). In that study, individuals who knew the domain very well were not troubled by unclear conceptual models because they could use their prior knowledge to understand what the model was intended to show. In this study, we extend that work by investigating the impact of prior domain knowledge in more detail.

Our research question, therefore, is: "how should roles be shown in conceptual models and does this depend on readers' familiarity with the domain shown in the model?" The next section discusses the concept of a role and the effect of domain familiarity, in turn.

2 THEORY

2.1 The Concept of a "Role"

Roles have been termed the key constituents of organizational structure (Walsh and Ungson 1991). Given their importance, it would be useful if analysts had a way to model them in conceptual models. Some conceptual modelling methods such as object-role modelling (Falkenberg 1976) and the business process modelling notation (BPMN 2004) enable analysts to model some aspects of roles. At present, however, no conceptual modelling method exists that fully supports this goal. In this paper, we describe guidelines for modelling roles using the syntax of an existing conceptual modelling method: the extended entity relationship (EER) grammar. The EER is an extension of the original ER model that uses the original concepts of entities and relationships and offers additional concepts such as subset entity types, parts, and wholes that are useful in modelling (Teory et al. 1986). We follow this approach because the ER model is the most widely used conceptual modelling method in practice (Davies et al. 2006). Therefore, if our guidelines prove useful, many individuals should be able to take advantage of them.

In past literature, researchers have associated the concept of a role with two other concepts: *relationships* and *interactions*. For example, Sowa (2000) and Guarino (1992) suggest that roles imply relationships among objects and that the role of an object is associated with properties the object acquires through relationships with other objects. Zhu et al. (2006) define a role as part of an object's behaviour (or responsibility) that is determined by the interactions the object engages in. Finally, Boella et al. (2007) suggest that roles are a relational concept associated with interactions. Building on past literature, we define roles in terms of interactions and classes. To help us do so, we draw on ontological theory. Ontology is a branch of philosophy that deals with the structure of reality (Angeles 1981; Smith 2001). Ontological theories have been used in IS research to clarify the meaning of real world concepts reflected in conceptual models (Wand et al. 1995). Because a role is a real world concept, ontological theory should be useful for defining and describing it.

We use concepts from the ontological theory of Bunge (1977) that provides a set of high-level concepts to represent real world phenomena, such as thing, property, state, and interaction. Bunge's ontology suits our purpose because it is formal, comprehensive, and has been adapted and extended in IS research (Wand and Weber 1990). As Bunge's ontology explicitly defines 'classes' and 'interactions' we use this ontology to propose guidelines for modelling roles. In Bunge's (1977) ontology, a thing *acts* on another if the first affects the state changes in the other. Two things are said to *interact* or if at least one acts on the other. The action of one thing on another is manifested via a special kind of property known as a mutual property (Bunge 1977, p. 102). A *mutual property* is meaningful only in the context of two or more interacting things (Bunge, 1977), e.g. the "salary of an employee" is mutual to the employee person and an organization.

We use the notions of class and interaction to define a role.[1] A class is viewed as a set of instances that possess a set of common properties (Parsons and Wand 1997). A *role* reflects interactions of an instance of a class with some other instances (in the same or in another class). Hence, a role can be defined in terms of mutual properties acquired by some instances of a class when they engage in interactions. We term the original class a *base class*. The instances of a role are instances of a base class that have additional mutual properties reflecting interactions. These instances form a subclass of the base class. It follows that a modeller can show in a conceptual model the differences between instances of the base class which do not assume the role and instances of the role by showing an additional set of mutual properties reflecting the interactions. For example, a "student" role can be modelled by mutual properties that arise between instances of the classes 'Person' and 'University.' Examples for such mutual properties are 'Date of Admission' and 'Tuition Fee.' Later we will show how this example would be modelled in EER using our guidelines.

2.2 Guidelines for Modelling Roles

Using the concepts defined above (role, base class, and interaction), we propose a set of guidelines for modelling roles in EER. These guidelines allow modellers to show roles in EER models using three constructs in the EER syntax: entity types, relationships, and subset entities.[2] Table 1 presents the guidelines together with examples from the student admission scenario described above.

No.	Guidelines	Example
1	Identify entity types whose instances might interact. These will be the base entity types.	'People' and 'University'
2	Identify and model the attributes of the base (original) entity types.	'Name' and 'Address' are attributes of 'Person' and of 'University'
3	Model the role as a subset entity of the interacting entity types. The attributes of the role should not include the properties of the base entity type.	'Student' entity type is a subset entity of 'Person' entity type

[1] In this paper we provide only the main principles to define roles. Full details are included in Bera et al. (2009)
[2] We use the EER syntax as mentioned by Teory et al. (1986)

4	Model the interaction as a relationship between the role entity type and the other entity type.	'Student' entity type has a relationship 'Admission' with a 'University' entity type.
5	Model the mutual properties that reflect the interaction as attributes of the relationship.	'Admission' has the attributes 'Student Number,' 'Date of Admission,' and 'Tuition Fee.'

Table 1: Modelling guidelines for roles

Figure 1 shows an EER model of the admission scenario described above that follows the proposed guidelines. It shows that a student is a role of a person. It also shows that when persons serve in the role of a student, they *interact* with a University, for example, by being admitted in it. Finally, it shows that several mutual properties emerge from the interaction (properties shared by the university and the student) such as student number and date of admission.

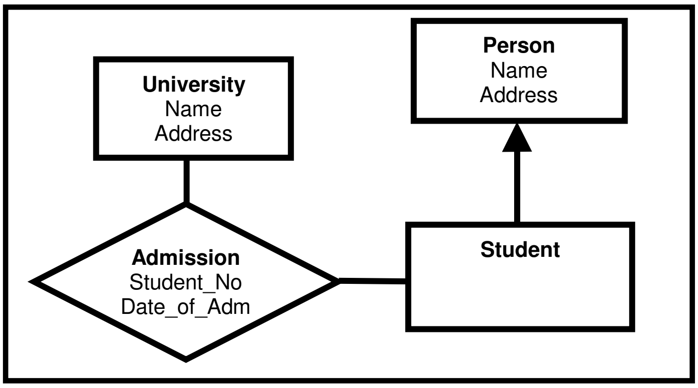

Figure 1. EER diagram showing that a student is a role of a person

In Figure 1, note that we used the symbol for a subset entity to show the role of the Student. This is appropriate because a role reflects a subset of entities that engage in interactions. However, one effect of our guidelines is that in a large conceptual model we might use the same symbol (a subset entity) to reflect *two* different types of phenomena – subset entities that reflect roles (e.g., a student is a role of a person) and subset entities that do not reflect roles (e.g., a man is a subtype of person). To ensure that readers are not confused by the same symbol being used in two ways, we suggest the following modelling convention: entity types that reflect roles will have *no* attributes whereas entity types that do not reflect roles will have *at least one* attribute. Since all properties of a role are either those of the original (base) class or mutual (included in the relationship), all properties of the instances of a role will be represented in the diagram.

2.3 Hypotheses

To develop the hypotheses, we first introduce two types of EER diagrams – *guided* and *unguided*. Guided diagrams are developed according to our guidelines in Table 1. Unguided EER diagrams violate guidelines 3 and 5. We focus on these two guidelines because they are particularly important for distinguishing roles from other kinds of subset entities, as noted immediately above.

To illustrate the difference between guided and unguided diagrams, consider the diagrams in Figure 2. These diagrams were created by violating the distinction between subset entities that reflect roles and those that do not. In the left side of Figure 2, the student role is shown in the relationship (Registered_In) rather than as an entity. Guideline 3 has been violated because no separate entity is created as a subset entity; rather the notion of role is embedded in the relationship between the two entities. In the right side of Figure 2, Student is shown as a subset of Person but Student contains attributes that are mutual to the Student and University thus violating guideline 5. The guided EER diagram of the same situation is shown in Figure 1 above.

Although the same domain concepts appear in both guided and unguided EER diagrams, we claim that because the unguided EER diagrams do not model roles explicitly, they do not provide a clear and accurate model of the domain. The guided EER diagrams show roles explicitly because a strict convention is used to show them (i.e., subset entities that have no attributes). We propose that such guided diagrams will be better understood than the unguided diagrams.

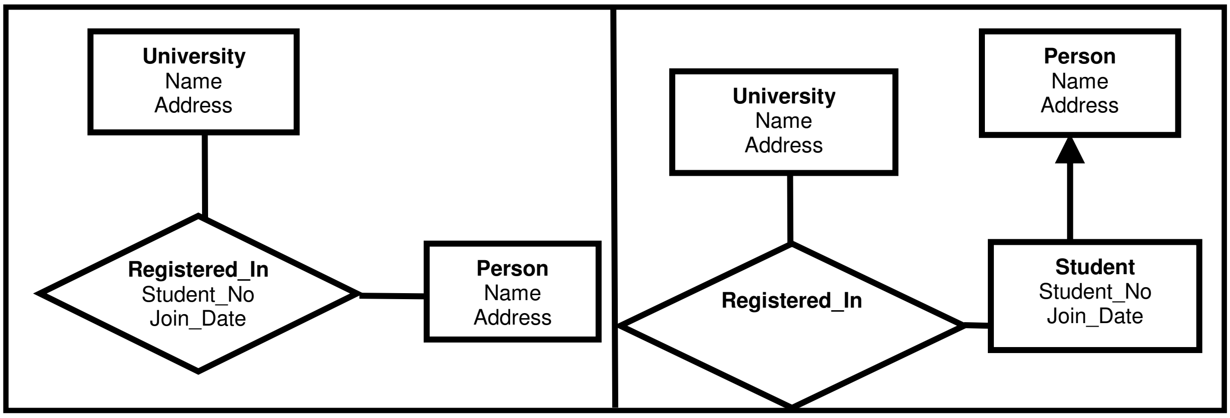

Figure 2: Two unguided EER diagrams - "a person is a student"

We base our hypotheses on the theory of cognitive fit (Vessey and Galletta 1991), which in turn is based on Newell and Simon's (1972) problem solving approach. According to this theory, when the type of information emphasized in the problem representation matches the type of information used in the task, the cognitive fit between the representation and the task enables the problem solver to develop a better understanding of the domain than when a match does not exist.

In the unguided EER diagram, the distinction between subset entities and roles is not explicit. This may result in ambiguity because two different constructs (roles and subsets) are represented by the same grammatical construct, a case of *construct overload* (Wand and Weber 1993). Because readers of such diagrams might not be able to resolve this ambiguity, the cognitive fit between the representation and the task is reduced, which in turn should result in lower domain understanding. Guided EER diagrams, on the other hand, clearly distinguish between subset entities and roles. They have no semantic ambiguities. Therefore when users refer to guided EER diagrams, the cognitive fit between the representation and the task is higher and should result in higher domain understanding.

Our argument just now did not consider the prior knowledge of the reader of the script. We suggest, however, that our argument should hold only when the reader has what we will refer to as a *moderate* understanding of the domain, i.e., the reader has a reasonable knowledge of it, but is neither completely familiar nor completely unfamiliar with it. If readers are completely familiar with the domain shown in a diagram and are asked questions about the diagram, they might answer the questions based on their prior knowledge rather than the diagram. Conversely, if users are completely unfamiliar with the domain shown in the diagram, they might not have the understanding necessary to utilize the diagram and answer questions (Freebody and Anderson 1983; Pretz et al. 2003). When domain familiarity is moderate, however, we predict that users will rely on the diagram to understand the domain. In past conceptual modelling research, conceptual models have been developed from domains that are moderately familiar to subjects, e.g., Gemino and Wand (2005) used a bus reservation domain and Burton Jones and Meso (2006) used a job application domain.

Overall, if individuals are given a guided EER diagram that reflects a domain that is very familiar to them, and are then tested on their understanding of the domain shown in the model, we expect that they would score highly on questions on problem solving, primarily because of their pre-existing knowledge. We expect that this would still be the case even if the individuals were given an unguided diagram of the domain. This is because these individuals can use their pre-existing knowledge to answer questions and can resolve semantic ambiguities in the script based on their prior knowledge. This argument follows from Ashcraft (1989) who mentions that contextual knowledge allows semantic ambiguities to be resolved. Likewise, Freebody and Anderson(1983) mention that individuals can interpret text successfully even if it is unclear if they have prior knowledge of the concepts in the text.

The situation is different, however, for individuals who are completely unfamiliar with the domain shown in a diagram. Users referring to unguided EER diagrams that show a domain that they are not familiar with, can no longer resolve semantic ambiguities in the diagram. Even if users are given a guided EER diagram, however, they may still not understand the domain shown in the diagram because they do not have enough prior knowledge with which to interpret the concepts in the diagram. Thus the advantage of the guided EER diagram over the unguided EER diagram is nullified when domain familiarity is very low (i.e., whether a diagram is semantically clear or semantically ambiguous does not matter). Therefore, we predict that when domain familiarity is very low, then the cognitive fit is also low whether users refer to guided or to unguided diagrams, resulting in low domain understanding. The above six cases are summarized in Table 2.

No.	Problem representation in EER	Domain Familiarity	Cognitive fit	Predicted Domain understanding
1	Guided EER diagram	High	High	High
2	Unguided EER diagram	High	High	High
3	Guided EER diagram	Moderate	High	High
4	Unguided EER diagram	Moderate	Low	Low
5	Guided EER diagram	Low	Low	Low
6	Unguided EER diagram	Low	Low	Low

Table 2. Effect of domain familiarity on domain understanding

Based on the above discussion we suggest the following three hypotheses:
H1: When the domain is very familiar, there will be no significant difference in understanding of the domain between individuals reading guided EER diagrams and individuals reading unguided EER diagrams.

H2: When the domain is moderately familiar, individuals reading EER diagrams will obtain a better understanding of the domain if their EER diagrams are guided than if they are unguided.
H3: When the domain is very unfamiliar, there will be no significant difference in understanding of the domain between individuals reading guided EER diagrams and individuals reading unguided EER diagrams.

In short, we expect an interaction effect between the guidelines and domain familiarity, such that our guidelines will have more benefit when domain familiarity is moderate. Because readers' mean level of understanding of each model should reflect their prior knowledge, we can plot this interaction as shown in Figure 3. We do not test this interaction effect statistically. Rather, we test the effect in each domain separately, as per our hypotheses above. Nevertheless, the figure illustrates the overall logic of our hypotheses. The strength of the effect (the angle of the slope) in Figure 3 is also stylized; our theory predicts an effect but we cannot predict a specific effect size *a priori*.

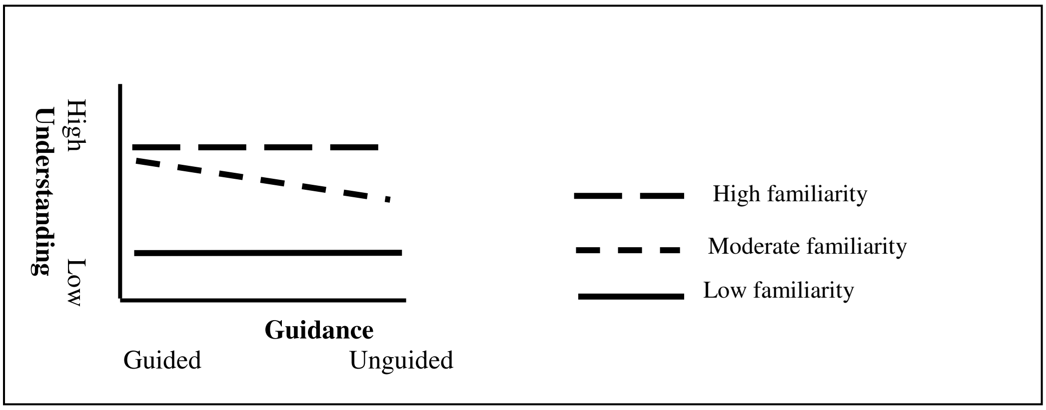

Figure 3: Interaction effect of guidance and domain familiarity on domain understanding

3 METHOD

3.1 Subjects and domain selection

A laboratory study was conducted to test the hypotheses with students enrolled in a post-graduate Accounting Information Systems course in a Canadian University as subjects. This sample was suitable for our study because all students in the course had completed an undergraduate degree and had several years work experience. Moreover, as the course was designed to 'fast-track' students into accounting jobs, it is reasonable to generalize our results to new recruits entering accounting roles in practice. 20 students participated in the pilot study and 73 participated in the main study. For participation, subjects received 2% credit for the course. For additional motivation, cash rewards of $20 were provided to participants based on their performance in the study.

A library domain, an aquarium management domain, and a pharmaceutical drug domain were used as the very familiar, moderately familiar, and very unfamiliar domains respectively. The library domain consisted of concepts such as Borrower and Library. Because our subjects had been university students, we expected that they would be very familiar with library concepts. Aquarium management was considered to be a domain where subjects had moderate familiarity. The domain has concepts such as quarantine, exhibited animals, and animal habitats. Although subjects might have visited an aquarium, we expected that they would not have detailed knowledge of how an aquarium is managed. Finally, for the very unfamiliar domain, we selected the case of a pharmaceutical drug--hydrocortisone. We did not expect students to know the effects of this drug. We recruited a Ph.D. student in pharmacology to write the drug narrative and help construct the EER diagrams for this domain.

3.2 Dependent Variable and Treatment

Our *dependent variable* was subjects' level of domain understanding. Prior research has shown that results from experiments can differ depending on the specific type of task that researchers use to measure understanding (Khatri et al., 2006). Following Gemino (1998), we measured domain understanding via inferential problem-solving questions. Such questions require subjects to develop creative solutions that go beyond the semantics explicitly shown in the diagrams. They have been argued to be a good measure of subjects' deep understanding of the meaning conveyed by a model (Gemino 1998). Appendix A includes sample problem-solving questions used in the experiment for the moderate and very low familiarity domains.

The *treatment* in the experiment was the use of a guided EER diagram versus an unguided EER diagram. Six EER diagrams were created, a guided and an unguided diagram for each of the three domains. As noted earlier, the difference between the guided and unguided diagrams was that the unguided diagram violated rules 3 and 5 of the guidelines, whereas the guided diagrams complied with all of the guidelines. For example, in the unguided aquarium management EER diagram (Figure 5, Appendix A), the entity denoting the role of Animal Handlers includes the mutual properties of Animal Handlers and the Aquarium. Similarly, in the unguided pharmaceutical drug EER diagram (Figure 5, Appendix A), the relationship indicating "bone formation by osteoporosis" includes the role of the "bone formation agent." Using the guidelines, bone formation agent is modeled as a role of the drug hydrocortisone (Figure 5, Appendix A). The main distinction between the guided and unguided EER diagrams of the pharmaceutical drug domain is that the former makes a distinction between roles and subset entities (e.g., Bone formation agent is a role but Hydrocortisone is a subset entity) whereas the latter does not make this distinction (e.g., Bone formation agent is embedded in a relationship). Similarly, in the aquarium management domain, the guided EER makes a distinction between subset entity types that represent roles and subset entity types that do not represent roles (e.g., Animal Handler is a role and has no attributes, while some other entity types are not roles and have attributes) whereas the unguided EER does not make this distinction (e.g., Animal Handler has attributes). Both guided and unguided EER diagrams have legends to facilitate subjects' understanding of the diagrams. To conserve space, only parts of the guided and unguided diagrams of moderate and very low familiarity domains are shown in Appendix A.

Because the experiment investigates whether a difference in the *semantics* of two diagrams affects readers' understanding of a domain (in this case, semantics about roles), it was important to control for non-semantic differences between the diagrams that might confound the results. One potential confound is the layout of the diagram (Shanks et al. 2008). We therefore ensured that there was minimal difference in the layout of the two diagrams of each domain.

3.3 Experimental design and procedures

A between-subject experimental design was developed. The subjects performed the experimental task for the three domains either using a guided version or an unguided version of the EER diagrams. The procedure was followed in two stages: training and experiment. During the training stage, subjects were given a pre-test questionnaire to identify their prior modelling knowledge and prior domain familiarity for each domain. Next, they spent 15 minutes on the basic concepts of EER diagrams. Subjects who later received guided diagrams also received additional information on roles. Subjects next practiced answering problem-solving questions using a simple case. Then they received feedback on their performance in the practice task.

In the experimental stage, each subject either received the guided version or the unguided version of the diagrams. First, the subjects were asked to describe the contents of the EER diagram as they understood them. This task was not used to obtain a dependent measure or control measure; instead, it was used to engage the subjects with the diagrams. Because problem solving tasks require deep understanding (Mayer 1983), we included this task to help subjects familiarize themselves with the diagrams before they did the main task. After they completed describing the contents of the diagrams, they performed the main task of answering the problem-solving questions. This sequence was repeated with the other two domains. The order of the domains was randomized. The same set of problem solving questions was used for both groups (guided and unguided EER diagrams).

To ensure that the differences between the groups can be attributed to the treatment, data on several control variables were collected: prior modelling knowledge, prior knowledge of the domain, order of the diagrams presented to subjects, and time to complete the problem solving tasks.

4 RESULTS

4.1 Descriptive statistics

A pilot study was conducted with 20 subjects to fine tune the measures used for the dependent variable. Slight modifications in the wording of the problem solving questions were made after the pilot study. The main study was conducted with 73 subjects. Two MIS PhD students were recruited as coders to identify the correct subjects' responses. A detailed coding document containing the possible answers for the problem solving questions was provided to these coders. The first coder coded all the responses where as the second coder coded the data for 30% of the subjects (randomly chosen). The reliability between the coders was high (alpha = 0.97).

The descriptive statistics are reported in Table 3. The mean domain familiarity scores were consistent over the three types of familiarity (3.41, 2.63, 1.50), indicating that the domains selected indeed were of varying familiarity. Also, the average problems solving score was much lower for the pharmaceutical drug domain (0.84) than the other domains (2.83, 2.92) as we expected.

Variables	Scale	Mean I	SD. I	Mean U	SD. U	Mean A	SD. A
1. EER modelling familiarity	1-7	3.49	1.40	3.32	1.18	3.40	1.29
2. Library domain familiarity	1-7	3.41	1.36	3.25	1.31	3.33	1.33
3. Aquarium mgt domain familiarity	1-7	2.63	0.97	2.12	0.87	2.36	0.95
4. Pharma drug domain familiarity	1-7	1.50	0.91	1.51	1.00	1.51	0.95
5. PS average – library domain	1-6*	2.99	0.99	2.68	1.30	2.83	1.17
6. PS average – aquarium mgt. domain	1-7*	3.31	1.04	2.55	1.24	2.92	1.20
7. PS average – Pharma drug domain	0-3*	1.04	0.98	0.65	0.85	0.84	0.93

P.S.: Problem Solving, **I**: Guided group, **U**: Unguided group; **A**: Average of the two groups, **SD.**: Standard Deviation. * The maximum range indicates the practical maximum score as obtained in the study.

Table 3: Descriptive statistics

Figure 4 graphically shows the effects of guidance and domain familiarity on domain understanding. The slopes of all three lines are negative. However, guidance has the strongest effect (highest negative slope) when domain familiarity is moderate, as we expected.

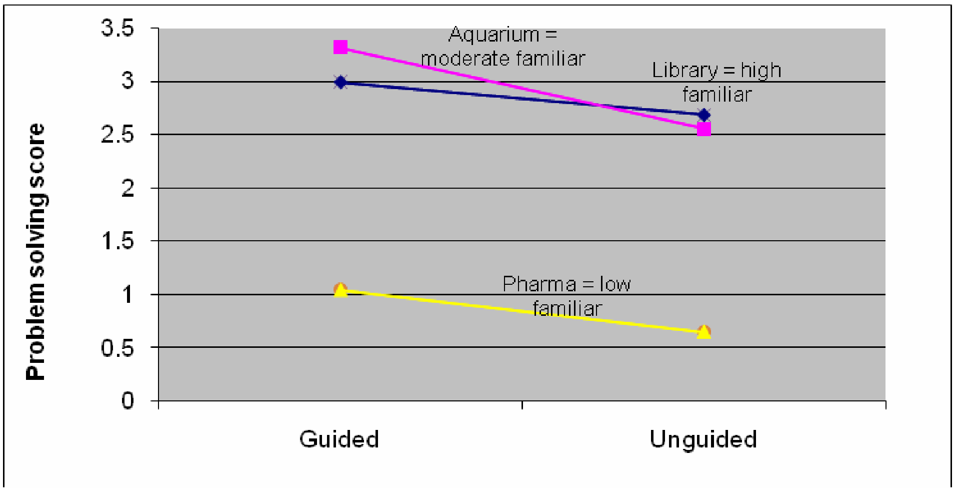

Figure 4: Interaction effect of guidance and domain familiarity on problem solving

Next we checked the reliability and validity of the instruments used for the problem solving questions (Table 4). The reliability of the problem-solving questions for the aquarium management and the library domains was 0.68 which was close to 0.7, the generally accepted value (Nunnally and Bernstein 1994). We also ran a factor analysis of the items used in the problem solving questions (reflecting the dependent variable problem solving). The results indicated the questions loaded distinctively on their own construct (displaying convergent and discriminant validity). On the basis of these tests, the reliability and validity of the problem solving items were considered satisfactory.

Domain	**Reliability**	**Validity**
Library	Cronbach's alpha 0.68	Factor loadings: 0.74, 0.83, 0.80
Aquarium Management	Cronbach's alpha 0.68	Factor loadings: 0.83, 0.80, 0.72
Pharmaceutical Drug	Cronbach's alpha 0.73	Factor loadings: 0.80, 0.86, 0.80

Table 4: Instrument Validity and Reliability

4.2 Hypotheses testing

We used ANOVA to test the hypotheses (Table 5). The results indicate that the treatment was significant for the aquarium management domain but not significant for the library domain, consistent with our hypotheses H1 and H2. However, in the pharmaceutical drug domain the treatment was significant (albeit only when using a liberal alpha of 0.10), providing some evidence against H3.

Domain	**Degree of Freedoms**	**Mean Square**	**F**	**Sig**	**Hypothesis**
Library	(1, 71)	1.71	1.25	0.267	H1 supported
Aquarium Management	(1, 71)	10.57	7.99	0.006*	H2 supported
Pharmaceutical drug	(1, 71)	2.76	3.29	0.074**	H3 not supported

*Significant at 0.05 level (two-tailed), **Significant at 0.10 (two-tailed)

Table 5: ANOVA Analysis

To test the effect of control variables in the moderately familiar domain, we performed an analysis of covariance (ANCOVA). The results in Table 6 indicate that except for the treatment, none of the control variables affected the results.

Variables	**F**	**Sig. (1-tail)**
EER Modelling Knowledge	0.22	0.64
Aquarium management domain familiarity	0.47	0.50
Time for Problem solving questions	0.03	0.87
Order of the diagrams presented	0.10	0.75
Treatment	5.22	0.03
Adjusted R^2 = 0.03		

Table 6: ANCOVA Analysis for Aquarium Management Domain

5 CONCLUSION

In this study we examined two issues in conceptual modeling: how to model roles and whether readers' prior familiarity with the domain shown in a model would affect the usefulness of guidelines for modeling roles. This research suggests that conceptual models can be developed to show roles more clearly but that the benefit of doing so depends on readers' familiarity with the domain shown in the model. It appears that our guidelines will be most useful when readers have moderate knowledge of the domain shown in the model. When readers are very familiar with the domain shown in the model, our guidelines do not seem to have much benefit. When readers have very little knowledge of the domain, the guidelines do help, but only a little. This result was contrary to our expectations but we should note that very low domain familiarity is not the same as *no* familiarity. If our study was done as in Parsons (2003), in which users were shown domains that they could not possibly have any knowledge of, then our results may have been more in line with our predictions. Nevertheless, this unexpected finding is promising because it implies that clear conceptual models can help individuals understand domains that they are very unfamiliar with.

Our study makes contributions to both research and practice. For research, we demonstrated a way to model roles in conceptual models and we extended past research on domain familiarity by showing that the effects of prior knowledge occur most strongly when readers have a moderate knowledge of the domain shown in a model (rather than a very high or very low knowledge of the domain). For practice, we provided a set of guidelines for modeling roles that practitioners can use when creating EER diagrams. Moreover, our results clarify for practitioners when it would be most useful to follow these guidelines.

More empirical research is certainly needed to corroborate our findings. For example, the study can be replicated using different constructs (such as composition) and modeling languages (such as UML). Studies could also be conducted to throw light on the cognitive processes that a user undergoes when exposed to diagrams developed on domains of different levels of familiarity. A process tracing study could be conducted for such a purpose. Process tracing might also reveal how the use of roles helps users to understand a domain. Future research could also focus on the effect of domain familiarity on different types of tasks other than problem solving.

References

Angeles, P. (1981), Dictionary of Philosophy. New York, NY: Harper Perennial.

Ashcraft, M. H. (1989), Human Memory & Cognition. Scott, Glenview, IL: Foresman and Company.

Bera, P., Burton-Jones, A., and Wand, Y. (2009) "Adding Organizational Semantics to EER: The Role Case." Working paper.

Boella, G., L. Torre, and H. Verhagen (2007), "Roles: An Interdisciplinary Perspective," Applied Ontology, 2, 81-88.

BPMN (2004), "Business Process Modeling Notation," BPMI.org.

Bunge, M. (1977), Ontology 1: The Furniture of the World: D. Reidel Publishing.

Burton-Jones, A. and P. Meso (2006), "Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis," Information Systems Research, 17 (1), 38-60.
Burton-Jones, A. and R. Weber (1999), "Understanding relationships with attributes in entity-relationship diagrams," in ICIS. Brisbane.
Davies, Islay, P. Green, M. Rosemann, M. Indulska, and S. Gallo (2006), "How do practitioners use conceptual modeling in practice?" Data and Knowledge Engineering, 58 (3), 358-80
Evermann, J. and Y. Wand (2006), "Ontological Modelling Rules for UML: An Empirical Assessment," Journal of Computer Information Systems, 47 (1).
Falkenberg, E. (1976), "Concepts for modeling Information Systems," in IFIP conference on modeling in database management systems, G.M. Nijssen (Ed.). North Holland, Amsterdam.
Freebody, P. and R.C. Anderson (1983), "Effects of Vocabulary Difficulty, Text Cohesion, and Schema Availability on Reading Comprehension" Reading Research Quarterly, 18 (3), 277-94.
Gemino, A. (1998), "Comparing Object Oriented with Structured Analysis Techniques in Conceptual Modeling (PhD Thesis)," University of British Columbia.
Gemino, A. and Y. Wand (2005), "Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties," Data and Knowledge Engineering, 55, 301-26.
Guarino, N. (1992), "Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases," Data and Knowledge Engineering, 8 (2), 249-61.
Hadar, I. and P. Soffer (2006), "Variations in conceptual modeling: Classification and ontological analysis," Journal of the Association for Information Systems, 7 (8), 568-92.
Hoffer, J.A., J. F. George, and J.S. Valacich (2007), Modern Systems analysis and design (4 ed.): Prentice Hall.
Khatri, V., I. Vessey, V. Ramesh, P. Clay, and S. Park (2006), "Understanding Conceptual Schemas: Exploring the Role of Application and IS Domain Knowledge," Information Systems Research, 17 (1), 81-99.
Mayer, R.E. (1983), Thinking, problem solving, cognition: W.H. Freeman and Company.
Newell, A. and H. A. Simon (1972), Human Problem Solving. Englewood Cliffs, NJ: Prentice Hall.
Nunnally, Jum and Ira Bernstein (1994), Psychometric Theory (3 ed.). New York: McGraw Hill.
Parsons, J. and Y. Wand (1997), "Choosing Classes in Conceptual Modeling," Communications of the ACM, 40 (6), 63-69.
Parsons, J. 2003. "Effects of Local Versus Global Schema Diagrams on Verification and Communication in Conceptual Data Modeling." Journal of Mgmt. Inf. Systems, 19(3), 155-184
Pretz, J. E., A. J. Naples, and R. J. Sternberg (2003), Recognizing, Defining and Represeting Problems. Cambridge: Cambridge University Press.
Shanks, Graeme, Elizabeth Tansley, Jasmina Nuredini, Daniel Tobin, and Ron Weber (2008), "Representing Part-Whole Relations in Conceptual Modeling: An Empirical Evaluation
" MIS Quarterly, 32 (3), 553-73.
Smith, B. (2001), "Ontology and Information Systems," (accessed November 11, 2008), [available at http://ontology.buffalo.edu/ontology%28PIC%29.pdf].
Sowa, J.F. (2000), Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove, CA: Brooks Cole Publishing Co.
Teory, Toby J., Dongqing Yang, and James P. Fry (1986), "A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model," ACM Computing Surveys, 18 (2), 197-222.
Vessey, I. and D. Galletta (1991), "Cognitive Fit: AN Empirical Study of Information Acquisition," Information Systems Research, 2 (1), 63-84.
Walsh, J.P.; and G.R. Ungson (1991), "Organizational Memory," Academy of management review, 16 (1), 57-91.
Wand, Y., D. E. Monarchi, J. Parsons, and C. Woo (1995), "Theoretical Foundations for Conceptual Modelling in Information Systems Development," Decision Support Systems, 15, 285-304.
Wand, Y. and R. Weber (1993), "On the Ontological Expressiveness of Information Systems Analysis and Design Grammars," Journal of Information Systems, 3, 217-37.
Wand, Y. and R. Weber (1990), "An Ontological Model of an Information System," IEEE transactions on software engineering, 16 (11), 1282-992.

Zhu, H., M. Zhou, and P. Seguin (2006), "Supporting Software Development with Roles," IEEE Transactions on Man and Cybernetics - Part A: Systems and Humans, 36 (6), 1108-22.

Appendix A: Experimental Materials

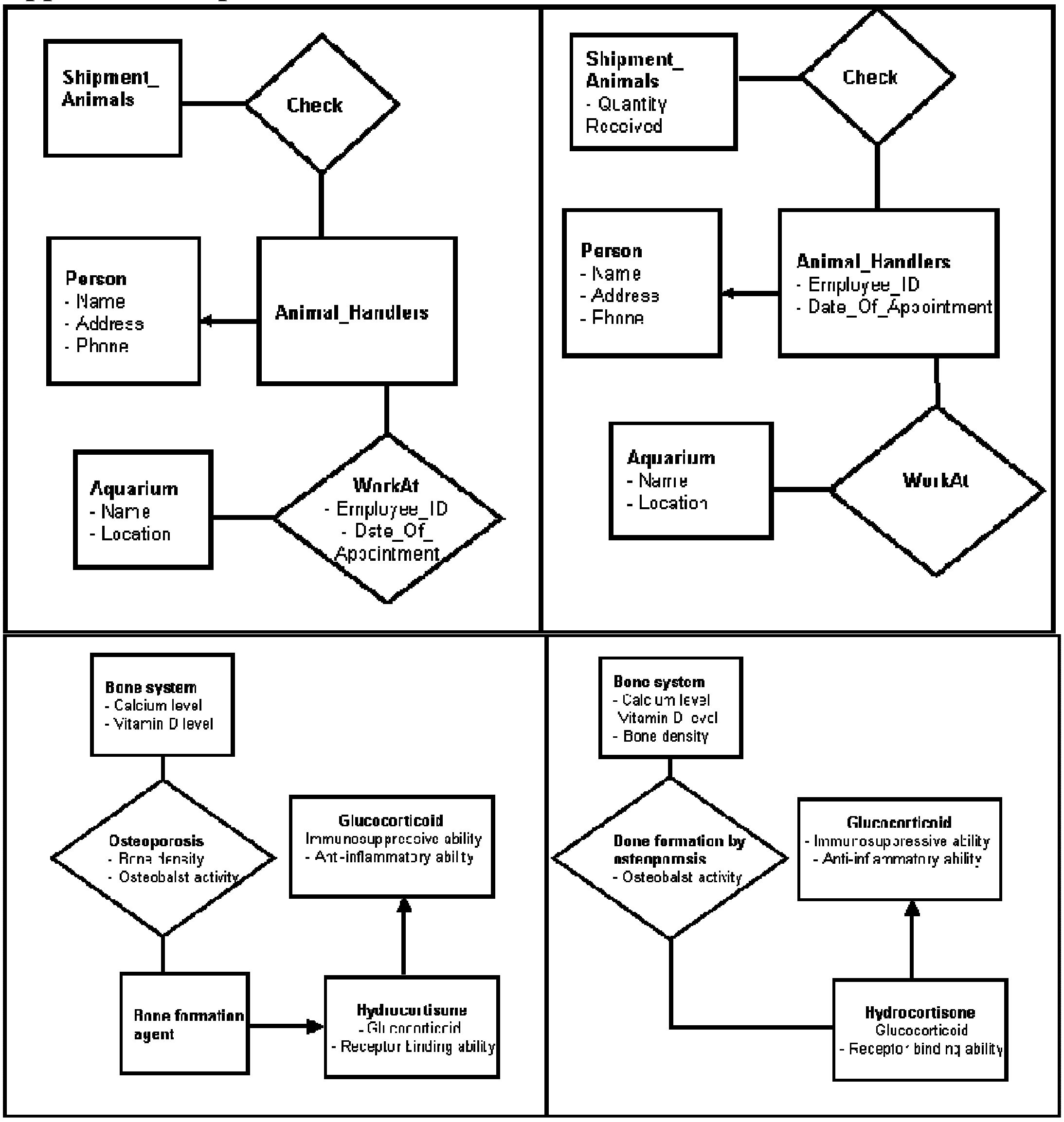

Figure 5: Parts of the guided (left side) and unguided (right side) EER diagrams from Aquarium management (top) and Pharmaceutical (bottom) domains

<u>Sample problem solving questions from aquarium management and pharmaceutical domains</u>

1. A new species of animal has just been discovered in Tasmania. The aquarium is thrilled to purchase this species. What measures should the aquarium take before it is displayed to the visitors?
2. Certain species of animals do not feed for a few days when they are brought in the aquarium. What measures will you suggest to make sure that these animals remain healthy?

1. A diabetic patient was administered Hydrocortisone intravenously. What possible reactions might happen?
2. What are the effects of hydrocortisone in an immuno-compromised patient with gastrointestinal problem?
3. An AIDS patient with pneumonia was given hydrocortisone. What are the possible effects?