Communications of the Association for Information Systems

Volume 11 Article 26

April 2003

An Interpretive Study of How Practitioners Use Entity-Relationship Modelling in a Ternary Relationship Situation

Steve Hitchman
The University of Melbourne, steve@infomod.fsbusiness.co.uk

Follow this and additional works at: https://aisel.aisnet.org/cais

Recommended Citation
Hitchman, Steve (2003) "An Interpretive Study of How Practitioners Use Entity-Relationship Modelling in a Ternary Relationship Situation," *Communications of the Association for Information Systems*: Vol. 11 , Article 26.
DOI: 10.17705/1CAIS.01126
Available at: https://aisel.aisnet.org/cais/vol11/iss1/26

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Communications of the Association for Information Systems

AN INTERPRETIVE STUDY OF HOW PRACTITIONERS USE ENTITY-RELATIONSHIP MODELLING IN A TERNARY RELATIONSHIP SITUATION

STEVE HITCHMAN
Department Of Information Systems
The University Of Melbourne
steve@infomod.fsbusiness.co.uk

ABSTRACT.

This paper discusses the use of interpretive research to learn about the practicality of entity-relationship modelling in ternary relationship situations. The findings can be compared with previous studies of novice modellers who used the same invented scenario. Previous research excluded the interaction between practitioners that would occur during a modelling session because novices were used to complete tasks in isolation. A team of experienced practitioners are shown to use entity-relationship modelling in a business context of social interaction about design. The interaction proves to be a key part of the modelling process. Practitioners 'talk with the notation' as well as using the notation to draw a diagram. The entity-relationship model constrains the social interaction because the model provides a way of talking about design. The practitioners use the model to talk about a normalized relational data structure in a way that undermines the idea of the entity-relationship model as an independent conceptual model. The findings show that theories from the field of linguistics explain why the model is used in this design-dependent way and suggest that this dependency may be inevitable. When the design conversation is about a normalized relational data structure there is no benefit to the practitioners from using a special notation for ternary relationships. On the contrary, the practitioner's design dependence seems to enable them to expose aspects of a domain that do not make business sense. The wider implication is that interpretive research's role is important in generating insights about the extent to which conceptual modelling is usable by practitioners. Interpretive research highlights the importance of being able to distinguish between ideas about conceptual modelling and ideas about how to apply modelling to practice.

Keywords: entity-relationship modelling, ERM, ternary relationships, normative language, interpretive research methods

I. INTRODUCTION

This paper discusses the use of interpretive research to learn about the practicality of entity-relationship modelling in ternary relationship situations. Conceptual modelling, generally entity-relationship modelling, is learnt by many Information Systems undergraduates as the basic theory underlying database design. This theory was first outlined over twenty-five years ago [Chen, 1976] and the central ideas about entity-types and relationships are still widely used today. Practitioners use the entity-relationship model to design databases using tools such as Oracle's Designer 6i and Computer Associates' ERwin.

Practitioners define their own ideas about how to use entity-relationship modelling. The practical definition of business relationships is an example of a practitioner definition ignored by researchers [Hitchman 2002]. A relationship in the ERwin tool, using the IDEF1X standard [IDEF1X 1993], is considerably different to the Barker [1989] standard for relationships used in the Designer 6i tool. Both ERwin and Designer 6i relationships are very different from the relationships proposed by Chen. Neither of these tools allows the designer to use the fundamental idea that relationships exist between any number of entity-types. Therefore, it should be interesting to find out what happens when entity-relationship modelling is used by practitioners to deal with ternary relationship situations. Asking whether aspects of a particular theory are practical, in the sense of suitable for use in a particular situation is an important question for an applied science.

Section II introduces ternary relationships, why they are important, and why their practicality is doubtful. Section III examines evidence about the use of conceptual modelling in practice. Section IV examines the interpretive research method and assumptions made. Section V examines the findings from the interpretive research. Section VI is a brief conclusion. Appendix I lists the scenario extract used in the research. Appendix II lists the experience of the practitioners. Two accompanying files (Appendices III and IV) present a copy of the transcript of the modelling session with some detailed interpretation notes and a Powerpoint presentation file that replays the diagramming actions of the practitioners.

II. TERNARY RELATIONSHIPS IN THE ENTITY-RELATIONSHIP MODEL.

> *"Conceptual design has long been recognized as the most crucial phase of the database design process … to obtain a system-independent global view …" [Dey et al. 1999, pp.453-454].*

The idea of 'system-independent' is that conceptual modelling takes place outside of the context of any existing technical solution or particular database management system (DBMS), for example. System-independence is the basis for proposing the entity-relationship model as a conceptual model in it's own right.

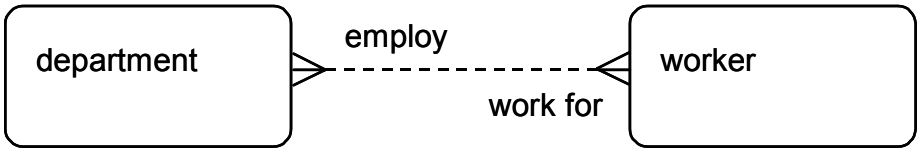
> *"A conceptual model is a language that is used to describe conceptual schemas. … A conceptual schema is a high-level description of the structure of the database, independent of the particular DBMS" Batini et al. [1992, p.6].*

Most authors differentiate the conceptual entity-relationship model from 'logical' models that are implemented by database management system (DBMS) products such as Oracle that partly implements the relational model. The logical model is the way that the users of the DBMS perceive the data structure, it is not the way that the data is stored. Chen [1976, p.10-11] defined the entity-relationship model in terms of

> *"Information concerning entities and relationships which exist in our minds … Let e denote an entity which exists in our minds". These … "conceptual objects in our minds." are represented as an "information structure … in which entities and relationships are represented by data." [Chen 1976, p.10,14].*

Relationships are central to the entity-relationship model and are defined to make the model system-independent. Since the original definition of the model, a relationship was always defined as an association among several entity-types (or an n-ary relationship). The simplest situation is a relationship between two entity-types, a binary relationship. Figure 1 shows a many-to-many

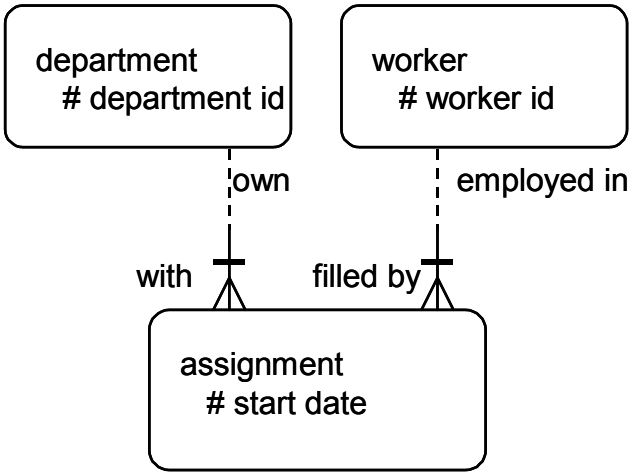
binary relationship between a worker and a department. The notation in Figure 1 specifies how to draw the diagram, but also specifies how to 'talk about' the domain. Two-way sentences (TWS) are formed for each relationship and 'say' what the diagram conveys to the reader. The way that a particular notation defines how to talk about the model is called the 'normative language' of the model. This normative language is not defined as a 'standard' part of the entity-relationship model but is in some versions of the notation used by practitioners.



A department may employ one or more workers
A worker may work for one or more departments

Figure 1. A Binary Relationship

Many-to-many relationships are not the simplest kind of binary relationships and are always decomposed in database design. Figure 2 shows the decomposition of the many-to-many relationship that results in a new entity-type. The two new binary relationships are simpler because they specify one-to-many situations. Details of unique identifiers are also included in Figure 2. The unique identifier of an assignment is a concatenation of 'department id', 'worker id' and 'start date'. The 'bars' across the relationships indicate identity dependence. This use of identity dependence is again found in some notations used by practitioners but is not part of the 'standard' model.



A department may own one or more assignments
An assignment must be with one and only one department
A worker may be employed in one or more assignments
An assignment must be filled by one and only one worker

Figure 2. Decomposition to a New Entity-Type

The decomposition in Figure 2 demonstrates a key practical problem with the entity-relationship model. It provides a choice of representing assignment as a relationship or as an entity-type. During design, for example, assignment may be proposed as an entity-type with the assignment relationship never proposed. This choice resulted in some critics of the entity-relationship model arguing that the model is flawed because it is not clear what constitutes its two main components. This unresolved argument dates back at least to Nijssen et al.. [1990]. Wand et al.. [1999] discuss the confusion surrounding the relationship definition, citing examples of the different representations of 'marriage' by different authors. "In short, the theory underlying the nature of and representation of relationships in conceptual modelling is unclear. … In our view, problems arise with relationships in conceptual modelling because their nature and underlying meaning are unclear." [Wand et al. 1999, pp. 495-496]. Current research includes attempts to use an ontology in order to resolve the issue. One way of looking at the issue is to think of the entity-type assignment as 'overloading' the model because an 'assignment' is a different 'sort of thing' compared to a 'worker' or a 'department'. The idea of overloading is that an entity-type is used to represent more than one sort of thing.

When the assignment is considered to be a relationship it is necessary to assign attributes to the relationship (data about the relationship). Figure 3 shows this idea using an adapted Barker notation. This notation is not widely used by data modellers and this is why the Barker notation must be modified to show the idea. This idea of relationships with attributes is central to the entity-relationship model.

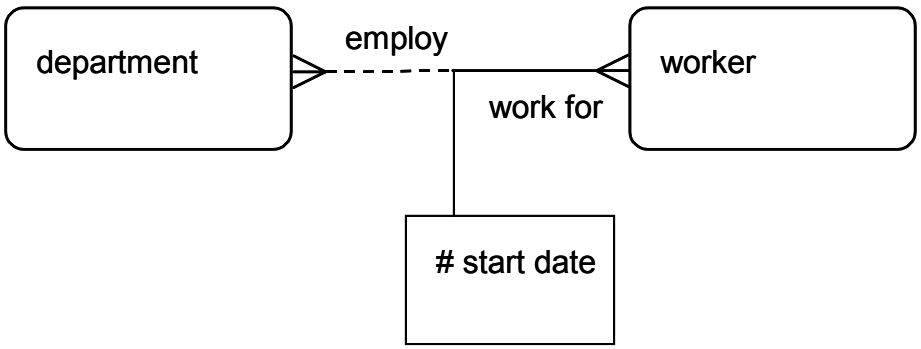


Figure 3. Attributed Relationship

A practical difficulty is deciding whether part of a domain should be represented by a relationship or by an entity-type, especially as both can own attributes. An interesting question is to ask why data model practitioners ignore relationships with attributes. For example, relationships with attributes are not available in either Designer 6i or in ERwin. Practitioners have no trouble deciding what is a relationship and what is an entity-type because attributes must be assigned to entity-types. Some design tools included a complication, though. In ERwin, but not in Desiger 6i, the model can be 'layered' with 'transforms'. With transforms, a decision can be made to leave some many-to-many relationships undecomposed (and unspecified) in one layer. A transform can be specified that decomposes and specifies the new entity-type in another layer. Using the same conceptual model it is therefore possible to create related layers with different levels of detail. In practice, decisions are made about what to leave undecomposed and when and to whom to expose the decompostion. Layering is not part of the entity-relationship model but seems to be related to the use of the model.

The basic relationship definition involves any number of entity-types. Users of the system-independent, conceptual model need to understand how to discover higher order relationships involving more than two entity-types. Relationships with three entity-types, called ternary relationships, are considered in all texts that describe entity-relationship modelling in detail. Figure 4 shows an example that was used several times in empirical research. The example is shown using UML (Universal Modelling Language) because it is not possible to show n-ary relationships in any of the notations commonly used by practitioner data modelers. In UML, a rectangle represents a class (an entity-type in this case) and the diamond symbol represents the ternary relationship. A '0..1' cardinality means 'zero or one', '1..*' means 'one or more' and '1' means 'one only'. This example is the expected answer of the inventors of the scenario in Appendix I. Using UML guidance [OMG 1999, p.3-73] we should use each pair in the ternary relationship and specify the cardinality of the other – "The multiplicity of a role represents the potential number of instance tuples in the association when the other two values are fixed."

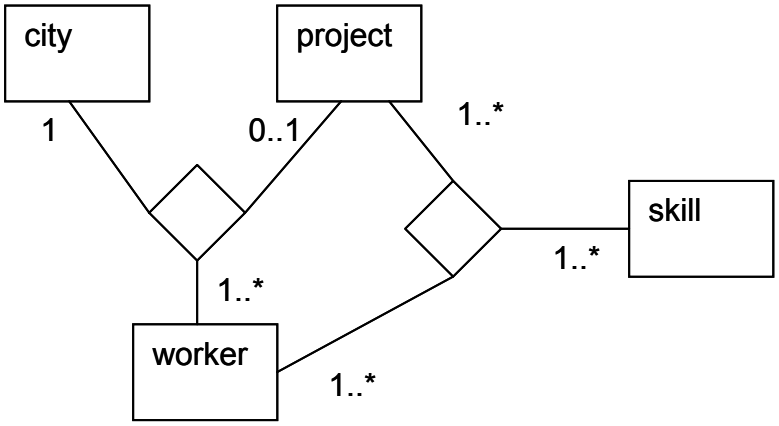


Figure 4. The UML Notation For Ternary Relationships

The UML notation specifies that, for example, a worker called Fred, when in Paris, is assigned to zero or more projects. Similarly, Fred on project Elephant must be in only one city, perhaps Paris. Fred could also be working on project Zebra in Dublin. Project Elephant in Paris is associated with one or more (i.e. must be associated with) Workers. This restriction implies that a project cannot exist unless someone is working on it. In theory any number of entity-types can be connected in this way, so this is the idea of an n-ary relationship. Curiously, UML offers no advice on using more than ternary relationships. The explanation of how to read ternary relationships is not standard. For example, Batini et al. [1992, p.23 and p.33] do not provide the same view of cardinality with regard to a pairing. Instead the cardinality is based on participation in the relationship as a whole. So the cardinality of a city would be zero (or possibly 1) or more – a city may be involved many times in the relationship. These different views of cardinality make a fundamental difference to the meaning of the relationship.

The main argument against ternary notation is that their use in data modelling actually undermines the design process [Hitchman 1999]. When practitioners are formally asked (this only happened once) about ternary relationship notation, the notation is perceived as problematic [Hitchman 2000]. The argued lack of practicality results from two causes:

Firstly, like the simpler binary many-to-many relationship, during database design the ternary relationship will decompose into a new entity-type. In the example in Figure 4 there would be two new entity-types, another kind of assignment (a worker on a project with particular skills) and yet another assignment (of a worker on a project in a particular city). Practitioners and their tools always used decomposition to deal with this situation; it is not possible to layer n-ary relationships. When both of these ternary relationships decompose into entity-types, problems are obvious to a data designer. For example, details of worker skills are repeated over projects and particular skills may disappear with projects. These problems are due to fourth and fifth normal form issues. Ternary relationships are thought to undermine the modelling process because they discourage the specification of required data, such as worker skills and because the business requirement is not normalized [Hitchman 1999].

Secondly, in practice there will always be one or more important interactions between the pairs of entity-types involved (this is also a reflection of normalization). In the example it is important to know a worker's skills, outside of any assignment to a project, otherwise how do we know who to assign with what skills? However, as soon as we obtain a list of worker skills, then it is these skills that will be assigned to a project. More likely we will also need a list of what skills are required on a project to be matched to worker skills. In other words, as soon as one or more of these many-to-many relationships are specified then the ternary relationship disappears.

Ternary relationships enjoy some prominence in the literature compared to higher order relationships that are not specially named. This prominence mirrors the special treatment of ternary relationships in UML. Ternary relationships are important because they represent the idea that the entity-relationship model is a conceptual, system-independent model based on n-ary relationships. One example of a paper describing higher order relationships is Dey et al. [1999], who define relationships using the same cardinality idea as Batini et al. [1992]. An example n-ary relationship is used involving a patient, doctor, drug and prescription. A prescription has a cardinality of one (for any prescription there can only be one instance), whereas a doctor has a cardinality of zero or more (a particular doctor can be involved any number of times in the relationship). Dey et al. [ibid, p.456-457] use a table to show:

"some sample instances of the relationship and the participating entities …

DOCTOR	*PATIENT*	*PRESCRIPTION*	*DRUG*
Davis	*Porter*	*prx01*	*Disperdol*
Davis	*Phang*	*prx02*	*Drefludan*

…"

This example illustrates three points about the relationship data. Firstly, the difficulty in distinguishing between entity-types and relationships is apparent from the description of these table rows as entities. Entities are otherwise represented by entity-types. Secondly, the data shown is only about the relationships with the entity-types. In a business situation there would also be data such as a start date and time (or timestamp). If the events are not contiguous then an end date and time may be required. A combination of timestamp and the four inherited keys is a candidate unique key for the table. Although there is no evidence from practice to support the claim, it is likely that these relationships will always require data such as event timestamps in a business situation. It is clear from this example that we would want to know when the prescription was issued, for example. Thirdly, the fact that the authors explain the relationship by using a table has a significance that is revealed later in the interpretive findings.

No established example in the literature describes a ternary or higher order relationship, from a real business situation that is sufficient to specify a real business database requirement. A closer examination of the prescription example, above, shows that the timestamp attribute seems to apply to the prescription, that is mandatory and unique. There will only be one 'ternary entity' for

each prescription. Could this relationship really be the prescription, issued by a doctor, for a patient, using a particular drug? What happens if two drugs are issued on one prescription? In other words, when the situation is placed in a real context this simple example becomes difficult to interpret.

Although n-ary relationships are fundamental to the model, no evidence from the practice situation describes what happens when modellers find a 'ternary situation', for example. Presumably, only binary relationships are used and many-to-many relationships, when discovered, simply flag a decomposition. If researchers invent a conceptual domain involving ternary relationships, then how will practitioners deal with this ?

III. THE EVIDENCE FOR THE USE OF RELATIONSHIPS IN PRACTICE

Wand and Weber [2002] set a wide ranging agenda for research into conceptual modelling, showing that little research had been undertaken on many important aspects. Their agenda includes understanding better how to use notations in particular contexts, the effects of underlying values and beliefs, and the fit between notations and different perspectives. Batra and Marakas [1995, p.190] surmised that the

> *"… academic community … have failed … since they have predominately focused on more expressive models … researchers have not attempted to conduct case or field studies … it is easy to observe that there are indeed wide differences between the academic and the practitioner focus on conceptual data modelling.".*

Russo and Wynekoop [1997, p.56-57] concluded that

> *" existing SDM (systems development methodologies) research demonstrated a reliance on normative research, largely focusing on the publication of conceptual papers and a paucity of empirical research addressing the use or efficacy of SDMs in practice … There is little evidence that existing SDMs are thoroughly evaluated before they are moderated or new ones developed. … it is dangerous to use an unevaluated SDM … Clearly, the new research paradigm must break out of the positivist box … interpretivist methods are desirable."*

The body of empirical research on conceptual modelling is large. Topi and Ramesh [2002] surveyed twenty-seven papers that evaluated some aspect of the usability of conceptual data modelling empirically over a twenty-four year period. Almost all of this research involved laboratory experiments with novices, generally undergraduates. Therefore this research reveals a lot about novice users and how undergraduate students deal with conceptual modelling in laboratory situations. The research is, of course, internally consistent, often using significance testing of statistical results. It is very difficult to know whether these experiments are externally consistent with real world modelling practice [Hitchman 1997, 1999]. Several research methodology reasons underly this state of affairs. For example, little research was done with experts in a practice situation. It is difficult to know how students and novices working in laboratory experiments can represent the experience and expertise of practitioners working in a real situation.

Specific issues concerning the experimental tasks set make it difficult to relate them to what happens in practice. A good example of this problem is the elegant experiment of Siau et al. [1996] discussed in Hitchman [1997]. In this experiment, graduate students were asked to decide if a diagram showing that a 'shareholder' may own 'shares' was sensible. Obviously it should make more sense to the graduate students if shareholders must own shares, otherwise they wouldn't be shareholders. However, for people in practice the optional relationship (may own shares) implies a completely different meaning. To them, the optionality means that, in a particular business situation, we may not currently know what shares are owned. For example, when we first know about a shareholder, do we also know what shares are owned? This

distinction is the difference between knowing that shares are owned and knowing what shares are owned. To a practitioner, it might be quite reasonable to say that a shareholder may own shares, in the sense of 'which shares'. Although the use of 'shareholder' as the name for an entity-type is rather suspect, shareholder is really a relationship name. So it is difficult to know whether the findings from even very constrained laboratory experiments are generally going to apply in the practice situation. The applicability of empirical research into Object-Oriented Systems Development [Johnson, 2002] is also in doubt because of a similar reliance on students to represent practitioners. This situation may be common.

A small body of research does reveal something about practice. A few surveys deal with practitioner perceptions [Hitchman 1995, Hitchman 2000] about aspects of modelling. The findings point to perceived practical difficulties in using aspects of conceptual modelling, particularly with business users. A few research papers used experienced modellers in laboratory experiments [Batra and Davis 1992, Shanks et. al 1993, Chaiyasut and Shanks 1994 ,Shanks 1997]. These experiments attempted to use differing high level frameworks to understand something about how practitioners work. Even in some of these experiments it is not clear how the experience of the modelers relates to practice. Batra and Davis [1992] used five experts with an unspecified experience of modelling and it is not clear how the expert categorisation was made in relation to practitioners. One expert was a graduate student, one was a full-time university teacher, and two others were part–time university lecturers. Using a high level abstract framework Batra and Davis concluded that the experts focused on a holistic understanding of scenario narratives, using different process models compared to novices (undergraduates). Shanks [1993] found that novices produced simpler models that translated scenario nouns ('literal models') and that were less complete. Experts made more use of generalization. Chaiyasut and Shanks [1994] compared four practioners with novice modellers. The practitioners experience varied, but averaged seven years intensive data modelling experience. The findings showed that, unlike the novices, the practitioners developed a holistic understanding and were able to reuse generic models from previous experience. Shanks [1997] used eighteen expert practitioners with at least four years experience as specialist modellers who had built at least ten conceptual data models. Shanks found that when the experts built a diagram from a narrative the results were more correct, complete, innovative, and flexible than those of novices. These experiments confirm that findings about novices do not apply to experts. In summary, few empirical modelling experiments clearly allow direct generalization to practitioners. Comparison with the Topi and Ramesh's [2002] list shows a lack of practitioner involvement in the research process.

This author concludes that there is no detailed evidence about how practitioners use relationships or, in particular, what happens when practitioners deal with a ternary relationship situation. Previous researchers observed a single modeller drawing a diagram from a written scenario, whereas modelling usually takes place with a group of participants talking through a domain (it is difficult to be precise here since there is no empirical evidence about what practioners actually do). The group discussion that usually takes place in modelling sessions, and is probably central to the method, is excluded from previous experimental studies.

IV. THE RESEARCH METHOD AND ASSUMPTIONS

This interpretive research takes a step towards practice but does not leave the laboratory situation. Therefore, the interpretive method can still be assessed against previously used laboratory research methods. This section shows as much about the method as possible so that future improvements to the method may be made.

A team of three experienced practitioners were videoed modelling an invented scenario previously used several times in student experiments [e.g. Batra et al., 1990, Shoval and Frumermann, 1994, Shoval and Frumermann, 1997, Shoval 1997, Shoval and Shiran, 1997].

This paper is restricted to reporting the modelling of just one part of the scenario that involved two ternary relationships in the diagram solution (shown in Figure 4) of the researchers. One of the ternary relationship constraints contains a 'one' rule that makes it unusual. This scenario was previously used to show, for example, that students using notations with ternary relationship symbols are more successful in producing the expected answer. The scenario extract is in Appendix I.

General guidance on interpretive case studies is given, for example, by Walsham [1995], Shanks et al. [1998] and King [1995]. This interpretive study adopts various techniques to fit the situation. The work of Carter et al. [2001] was used to provide guidance on the use of videoing as a technique to elicit knowledge about practitioners. The Carter et al. [2001] technique uses an expert commentary on videoed events and in many ways is similar to this author's own interpretation of the videoed modelling session. The interpretation could be considered to be an expert commentary on what the practitioners did. The author himself has more data modelling experience than any of the practitioners who took part in the experiment, using the measures discussed below. However, the modelling situation was much more constrained than those considered by Carter et al. [2001], and as often happens with interpretive analysis, the techniques are modified to fit the situation. We offer a direct interpretation of events and do not aim to use any pre-defined theoretical framework.

One aspect of being able to generalize the findings (i.e. will these findings apply to other data modellers) is knowing some detail about the respondents' experience in the ERM method. The questions asked to elicit the three respondent's experience and their responses are shown in Appendix III. One of the respondents is clearly more expert, but all have experience of large models and of using the Barker notation. At the time of the research, all three practitioners worked at the same business site. The researcher was himself involved in working at the site as a data modeller for over six months, so the researcher knew the participants. This knowledge helped to gain an inside view of the way the practitioners worked. For example, the researcher knew that the practitioners did not know either of ternary relationship notation or of UML. The potential disadvantage of the researcher being perceived as having a direct personal stake in research activities was mitigated by the brief that was given to the practitioners. The practitioners were told that a scenario was developed by other researchers and was previously used in laboratory experiments with undergraduate students who were asked to model the domain described using different notations. The researcher was seen to be neutral in that he did not own the scenario task and was seeking to find out how experienced practitioners would model the same scenario. They also understood that no one previously studied practitioners at work on the scenario.

The scenario provides a positivist constraint – the domain is pre-defined and we can share and always know about what sentences the practitioners are trying to make sense of. Thus, we can always compare our own interpretations with those of the modelling team. The advantage of using this scenario is that we can directly compare the performance of the practitioners with both the invented solution of the previous researchers and with the findings made about the students who used the scenario. The drawback of using a scenario is that it represents an artificial modelling situation. Therefore, we are not able to know whether the practitioners would work in the same way with 'real users'. For example, the practitioners may themselves use normative modelling language pro-actively. Fortunately, as will be seen in the analysis, the modelling team themselves provided a clear answer about the relevance of the scenario to a real life situation. The scenario's published answer t can be appraised to see what was in the minds of the inventors. This published answer was shown in Figure 4 and discussed in the UML context. This situation is different to a modelling session with users where the modelled result cannot be checked against an answer. Therefore, using the scenario results in simpler interpretation of the situation.

An indication of the difficulty in using scenarios to represent 'real' modelling is reflected in the time taken to complete the experimental task compared to practice. English [1999, p.131] briefly describes a data modelling case study that involved 20 domain experts from five management levels and different business areas. This case study involved ten days of data modelling sessions over a five week period. The final model contained 110 entity-types with 512 attributes and consensus definitions. This output averages at 11 entity-types per day – given an eight hour day that's slightly over one entity-type an hour. Later, the model also underwent a substantial validation, walking through 300 information views and then finding one new entity-type and 30 new attributes, and two changed data relationships (which represents a very high level of initial model quality). The average time taken by students to complete the entire research scenario (with the equivalent of around 15 entity-types) was around an hour. The modelling team reported on here took around 33 minutes to model 5 entity-types. Clearly more goes on in a real life modelling session than occurs in transcribing a scenario. The practitioners took considerably longer than the students to try to complete the scenario, which seems counterintuitive as they should be able to use their experience to complete the task quicker.

The video was analysed directly and also by interpreting both a verbal transcription and the actions made in diagramming. A key aim was not to impose a pre-existing framework on the modelling process, but to interpret directly what a team of practitioners do when modelling from a written scenario. To be systematic about analyzing the modelling session, it was transcribed and the transcription was used several times to replay the modelling session so that what happened could be fully understood. To expose the basis for the interpretation two other files are available:

- The transcription is available in Appendix III (there are a few summarized sections) so that the reader can follow the events and the interpretation. This approach is similar to the 'editing' method described by King [1995]. The transcript may be of use to other researchers and may also be useful for teaching undergraduates how modellers work.
- In addition to the verbal transcript, 26 diagramming 'actions'[1] can also be tracked to show the stream of events during the session. It is possible for the reader of the transcript to re-create the diagram from the transcript and all of the diagramming actions can be replayed in the Powerpoint slide presentation in Appendix IV.

This interpretation attempts to uncover what happens when practitioners use entity-relationship modelling, it is not necessarily designed to build any theory or to develop concepts. The interpretation and analysis aim to provide rich insights into specific aspects of modelling.

V. SESSION INTERPRETATION

Much of the detail of the data modelling is left to appendices. Some conceptual modelling researchers will be more interested in the transcript details while others in checking the interpretation against the transcript. The transcript enables to reader to make their own judgments about whether the author's interpretation of events is reasonable. This ability on the part of the reader is an important part of the interpretive method and is similar to exposing the statistical basis of empirical research. This section is a summary of the interpretation and a discussion of the findings. The interpretation is organized into

- normative language,
- social construction and overloading,
- the 'I' symbol for unique key inheritance,

[1] Sometimes several particular elements were combined when they occurred in the same action

- ternary relationships,
- the validity of the scenario, and
- a discussion of whether the interpretive research was successful.

NORMATIVE LANGUAGE

The way that a particular notation defines how to talk about the model is called the 'normative language' of the model. Hitchman [2002] suggested that normative language would reveal a better understanding of modelling practice. The transcript clearly shows the use of a normative language. Although the practitioners use the Barker notation, the use is very informal. TWS are rarely used in a formal way. The practitioners adapt to the scenario language immediately using "many workers *can* work on a project", for example. It is not clear whether the informal use of 'can' or 'may be' reflects the use of a scenario or whether it is a reflection of practice. Normative language examples include:

> *"But a worker can work on many projects"*
>
> *"This now says that a worker can only be assigned to one of those rows (project location)"*
>
> *"No (agreeing) because there's a many there"*
>
> *"The worker can only be assigned to a city once"*
>
> *"Can be assigned to many cities"*
>
> *"This model will not enforce the rule that a worker can only work on one project in one city. This will allow the worker to work on many projects and these projects would all be in the same city"*
>
> *"A worker has a number of skills, does that mean a worker used a skill set on a particular project assignment? Or is it just a number of their skills are used on a project assignment .. that's suggesting .."*

The sentences are related to the diagram but diagramming is done informally. For example, relationships are left partly unspecified. The practitioners are positing simple, short sentences and checking that each sentence makes sense, although they do so in an informal, conversational way. Quite often, after a drawing activity, the practitioners read the diagram to check both that a specified sentence makes sense, and to look for further implications. These simple sentences are often informal statements of part of the relevant TWS. The practitioners are talking through the domain but using the normative language of the model. It is important to make the point that they are not translating from 'everyday English' but the normative language of the model is being used directly to constrain the conversation. This is why the normative language is important. Therefore modelling is based on talking through the domain and is not primarily a 'drawing' activity.

At different times the diagram reflects a version of what was agreed, but does not reflect many of the nuances of the discussion. Figure 5 shows the informality of the diagramming. 'Crow's feet' to specify cardinality and dashed lines to specify optionality are not shown. The diagram mirrors the fluidity of the discussion at this point. The team members take a lot for granted in reading the diagram. This diagram means very little outside of the transcript.

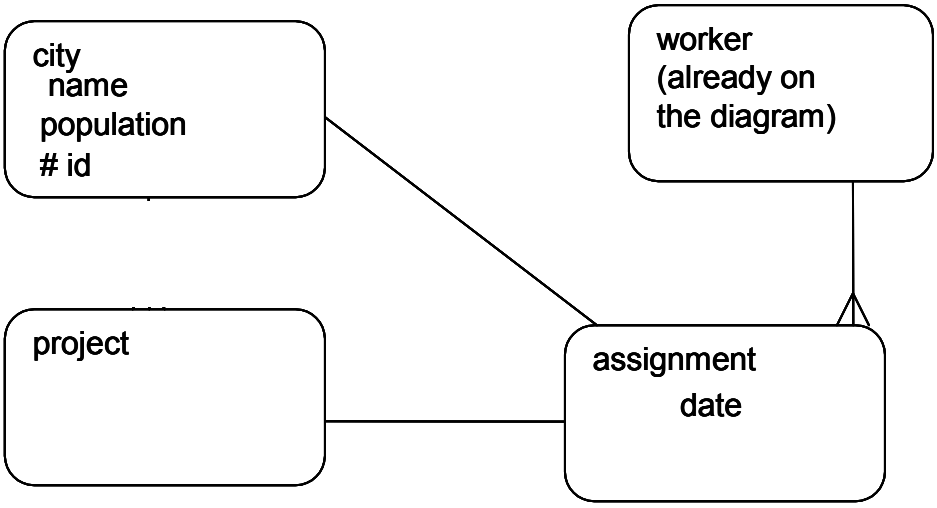


Figure 5. The diagram After Action 13.

Examples are often used to make sense of the situation. Evidence for this conclusion can be seen where the practitioners start to use examples to clarify entity-types and relationships:

> "*Right, so we could have another worker here (worker skill) who has a skill say accountancy, for example, supposing they are an accountant and an architect, they could be an accountant on one project an architect on another (pointing to assignment) and in there (worker skill) we would have one item of this would be them being an accountant on project one and another one with them being an architect on project two. What stops them being an architect on project one ?"*

SOCIAL CONSTRUCTION AND OVERLOADING

Overloading is the idea that an entity-type is being used to represent more than one sort of thing. Overloading is a reflection of the lack of agreement about what an entity-type actually represents in a conceptual model. The contention is that an overloaded model is difficult to use. Evidence of normative language sheds light on the issue of overloading entity-types. In the scenario, an assignment is probably the clearest case of something that could appear as an entity-type or as a relationship in the entity-relationship model. On the other hand a worker can only be an entity-type. Fred is a worker and we could all agree that Fred exists and we could go and shake hands with him. On the other hand, an assignment is a more nebulous concept. Searle [1995], discussing social construction, uses the idea of a 'brute fact' to explain why some concepts seem more solid that others. City is similarly solid (or brutish). We can walk around Paris, for example. Project and Skill are less solid, although we can all talk about a project to design and implement a database (project Elephant) or about a skill called data modelling. Assigning Fred to project Elephant is different again.

To understand the differences in these concepts we need to look outside of conceptual modelling theory [Veres and Hitchman 2002]. It is clear from the transcript that all of the entity-types are socially constructed. Fred may be a brute fact, but a worker type is socially constructed. The linguistic theory of Jackendoff [2002, p.308-309] explains what is really happening and seems to provide a missing theoretical basis for the modelling process:

> *"… we should properly think of the 'perceptual world ' not as absolute reality but as 'reality' constructed by our perceptual systems in response to whatever is 'really out*

> *there' … The perceptual systems … are not concerned with a 'true model of the world' in the logical sense, but with a 'world model' good enough to support the planning of actions …the perceived world is reality for us. … A percept is … constructed by the perceptual systems in response to stimulation from the outside world. Although a percept does not necessarily correspond exactly to what is 'actually out there' (especially with virtual objects), the experience that accompanies having a percept in one's … mind is that of an object in the world. … we are ultimately concerned with reality for us, the world in which we lead our lives.".*

This 'reality for us' is clear where the practitioners are talking about project-location entities as though they were rows in a table:

> *"This now says that a worker can only be assigned to one of those rows (project location) so a worker can only work in one city that the project is in."*

Modelling involves tacit knowledge about the working environment. A row in a table is as much a thing to these practitioners as a worker called Fred. Fred exists as a 'brute fact' but also 'exists' as a row in a table. Including tables and rows in the social construction is a reflection of the design context. The practitioners' social construction includes the tables required for a normalized data structure. Consequently, conceptual issues can be resolved by reference to data design, by understanding what tables would be required in a normalized structure. The language of rows and tables is hardly used at all in the transcript, but a designer can recognize that all of the entity-types proposed create a one-to-one mapping with relational tables in a design. In practice, then, the entity-relationship model is being used to design a set of normalized tables.

It is important to stress that this is not some kind of 'implementation specific' modelling that corrupted the entity-relationship model. Rather, the choice of entity-types takes place in the practitioner's context. Assignment, employees, and project-locations are all reality for the practitioners. The important issue is not whether entity-types are different sorts of concepts but whether the entity-types can be talked about in the same way so that everyone understands what is meant. This is consistent with Jackendoff's view of the use of different concepts in linguistics. Language is designed to manipulate different sorts of things on an equal footing, so overloading ceases to be an issue. Therefore the interpretive method reveals the social interaction and a theoretical foundation for understanding the 'overloading' issue that may not be apparent outside the practice situation. Although the situation is based on socially constructed facts this does not imply that the reality is socially constructed [Searle, 1995]. Using Searle's 'X counts as Y in C' explanation of social construction it is easy to see that, for example, a row in a table can count as something real to talk about. The conversation is still conceptual modelling because a row in a table is a construct of the relational model, it is not the physically stored data. However, this use is not system-independent. The practitioners are making sense of many-to-many and higher order relationships by grounding them in normalized table design.

The argument for the relational model (but not the entity-relationship model) always was that it is deliberately based on a well-understood business concept – a table (a relation). The rationale is that business users will be used to dealing with tables, for example that cross reference in the same way that assignment works. Business users will also be familiar with using screens that are views of the underlying data structure. Therefore, from the relational model point of view, it is reasonable to assume that business users will be able to understand an assignment or a 'project-location' because within the business context these constructions are normal. This use of tables

seems to be the practical solution of the problem of what constitutes a relationship or an entity-type when using the entity-relationship model.

This interpretive research does not provide evidence about whether business users are privileged to social constructions involving tables, because no business users were involved in the research. However, it is the social constructions rather than the notation that are revealed as the issue by the interpretive research. The extent of user social construction is an area for further research. Layering might be a reflection of the fact that some of the people from the business who will be involved in requirements specification are not privileged to the social construction of the tables in a relational database.

These findings raise interesting questions in the context of the generally held view that a

> *" … true conceptual data model should capture the essential characteristics of the domain of interest, and not necessarily the structure of the database."* Topi and Ramesh [2002, p.4].

In a sense, the practitioners have no choice about using their own context to talk through the domain. Indeed, Chen's original definition of entities as 'existing in our minds' emerges as a system-dependent idea when the model is used for data design. Therefore, these findings raise the question of whether the model can be system-independent in practice. Other questions for practice include:

- How can a business user help to specify data design requirements when they do not understand a normalized relational context? Should business users understand this kind of structure anyway, in order to understand business data?
- Is there any advantage in having entity-types that do not represent normalized tables when the object is data design? Are there some conceptual concepts that can be represented by entity-types that are not normalized tables? What would these concepts mean?
- If the entity-type is used to represent social constructions that are not normalized tables, is it a good idea to have a conceptual model where the key component can mean different things to different people? Do we need a completely different sort of model for these different social constructions?

THE 'I' SYMBOL FOR UNIQUE KEY INHERITANCE

The 'I' symbol to indicate an inherited unique identifier component is also strong evidence for the need to examine practice. This use of identifier inheritance seems to be very important. Figure 6 shows the diagram after action 21. After the final action (26) the diagram is the same but the 'I' on the project-assignment relationship is removed and the attribute 'time spent' is added to the entity-type worker skill. It will take the practitioners nearly half as long again to make these changes after action 21. This final (but essentially abandoned) diagram is unlike any model considered in the research literature, or any proposed and expected scenario answer. For example, understanding of 'assignment' is changed when the 'I' is later removed from the project-assignment relationship. This change creates an interesting 'ripple' effect on the meaning of worker skill because of the second inheritance of the worker-city identifier.

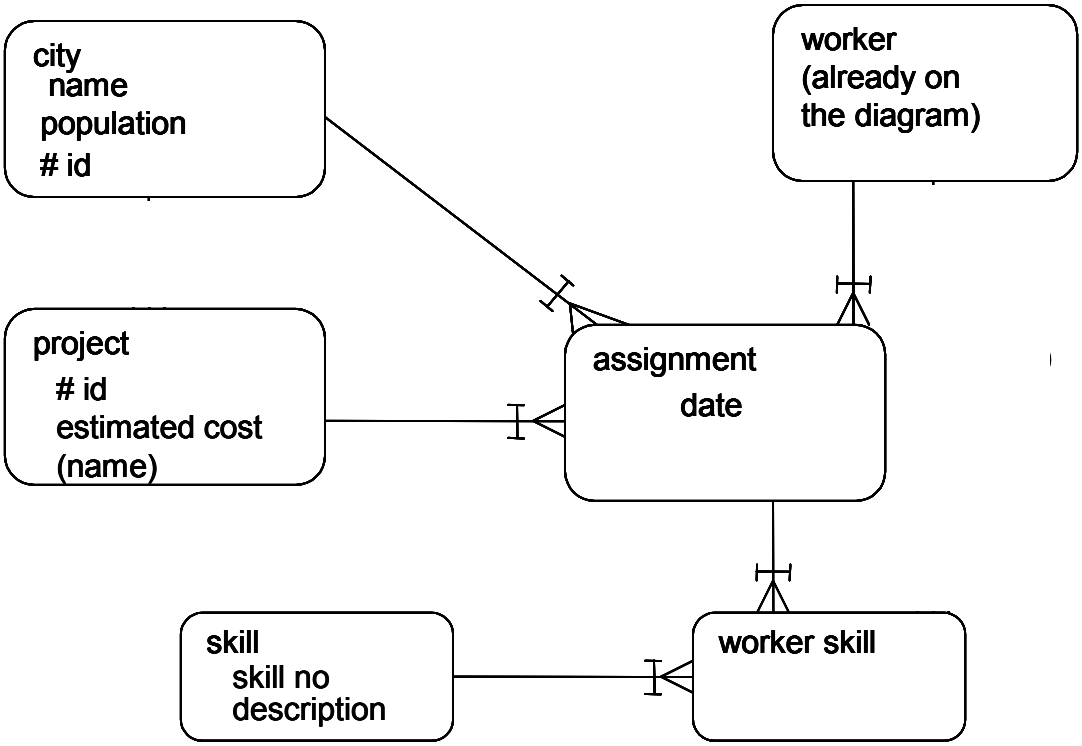


Figure 6. The Diagram at Action 21

It is worth noting that the IDEF1X notation used by the ERwin tool is predicated on recognizing unique key inheritance through 'identifying' and 'non-identifying' relationships. Therefore, it would be useful to know more about why this notation is used and what effect it has on the modelling process. In this example the removal of the 'I' notation from the project assignment relationship is an attempt to enforce the 'one' constraint between city and worker (because there can only be one unique value for any city-worker combination, say Fred in Paris). However, the perceived interaction between the 'I' notation and a mandatory relationship is unclear in reading the transcript. This area would benefit from further research.

TERNARY RELATIONSHIPS

A considerable amount of time is spent trying to understand 'can be assigned to only one project in a given city'. The underlying reason for this constraint is that although it seems to be a superficially simple idea, it does not 'make sense'. The scenario authors intended the sentence 'A worker can work on many projects, but can be assigned to only one project in a given city' to flag a ternary relationship between worker, project, and city. The practitioners use the previous (first) sentence 'many workers can work on a project', together with 'a worker can work on many projects' to derive a many to many relationship between worker and project *before* they consider the ternary information. The practitioners decompose to obtain the 'assignment' entity-type. It seems to be difficult for the practitioners to accept that an assignment is based on a project and (one) city. The assignment is to the project, the location is either part of the assignment decision process, or is simply accidental. The practitioners do not think of the location as a core part of the assignment.

Several strands of evidence undermine the relevance of n-ary relationships. Firstly the scenario verb 'assigned' became the 'assignment' entity-type. The practitioners talk about both assigning and assignment throughout the scenario, sometimes in the same sentence. For example:

> *"The worker can only be assigned to one project in one city so should the assignment be to the project ?"*

An assignment and a 'project skill' are real to these practitioners and so it makes sense to create them as entity-types rather than as relationships. Several examples from the transcript illustrate this point:

> "*So, you're now saying that assignment is current assignment"*
>
> *"So, this thing here (pointing to worker skill on the diagram) each one of these is one skill being used on one project"*
>
> *"No, isn't it just that these two things that we called assignment and project location, aren't these the same thing ?"*
>
> *"Something between worker-skill and assignment"*
>
> *"Well, you've posted your assignment down to worker skill"*

The name assignment is used for 'project-worker' but no other name is given for worker-skill. Even though worker-skill seems to be an overloading of the entity-type construct it makes no real difference in the conversation. There is no linguistic 'problem'.

Second, a lot of evidence in the transcript shows that the practitioners need to understand some of the binary relationships between the three entity-types involved in the ternary constraint. Some examples are:

> *"It doesn't say whether a project is wholly within a city"*
>
> *"Working on a project in a given city, you're saying that a project only takes place in a given city and doesn't span cities."*
>
> *"Yes, one project in a given city so does that mean that the worker works in the one city or that the project is in the one city?"*
>
> *"A worker has a number of skills, does that mean a worker used a skill set on a particular project assignment?"*
>
> *"If you are a new employee with seven skills on just one project only using two skills how do you know what the others are?"*

To make sense of the situation the practitioners need to understand whether projects span several cities, what skills are owned by workers, and what skills are needed for a project. Once 'assignment' or 'worker skill' are invented, and form a basis for understanding the situation, the need for a ternary relationship disappears.

The third piece of evidence concerns the problem posed by the 'one city' constraint. The practitioners struggle to understand this throughout the session and approach the constraint from various viewpoints. One of the strengths of the modelling process they use is that they will re-examine a disputed issue from several viewpoints in order to thoroughly talk through their ideas.

For example, what is the business logic in constraining a worker to one project in one city, but assigning them to work on, say, two projects in two different cities? Can a worker be assigned to the same project in two cities? Is this a constraint concerning workers and cities, regardless of project? The 'one' constraint would be unlikely to be embedded in a data structure anyway, being susceptible to business change. This conversation is a reflection of the practitioners' difficulty in viewing location as an integral part of assignment. Location might be a consideration in the decision about assignment, but it is most unlikely that a business would build in a fixed restriction about location. From a data structure point of view, it is difficult for the practitioners to see the problem as other than a worker-project assignment.

A final point to make concerns the advantage of using a simple notation. Suppose that a real life ternary relationship exists, for example outside of the need to understand associated binary relationships. We can see that, in practice, it makes sense to talk about a ternary relationship as an entity-type, like an assignment. When Dey et al. [1999] explained their invented prescription n-ary relationship, they did so through the use of a table. Dey's approach raises the question "why a special notation for this situation?"

Therefore, although the scenario builders used a situation that seems to contain 'real' ternary relationships, the practitioners never accept these relationshipships in business terms. Location is something to be considered during the process of assignment, but not something to be embedded in the data structure. The worker-skill-project ternary relationship is only understandable in the context of several decomposed binary relationships, such as worker-skill and project-skill. One might argue that the practitioners are not used to thinking 'ternary', or that ternary thinking is harder. The biggest problem with this point of view is the lack of evidence that ternary relationships 'exist' in a practical data modelling situation. We did not find any published examples, from a real-life situation, of a ternary notation that specifies a business database requirement. In this context it seems appropriate to interpret the findings as showing that, in a business situation, one does not need to think 'ternary'. The interpretive research cannot 'prove' this point, but does highlight the idea that researchers assume that situations are 'ternary' when they are not viewed that way by practitioners. Ternary relationship notation does not seem to be useful for data modelling, which could explain their lack of use in practice.

THE VALIDITY OF THE SCENARIO.

The normative language used by the practitioners is predicated on the idea that they are generally modelling 'over time' because otherwise some data is lost. The practitioners' key problem throughout the scenario is deciding whether they should sometimes adopt the implied (but not explicitly stated) contrary position that relationships will show just the current situation, a snapshot. A snapshot interpretation would imply that 'a worker may be assigned to one and only one city at the moment and we do not need to know whether they were there in the past'. This interpretation presents a real problem since 'A worker can work on many projects, but can be assigned to only one project in a given city' is definitely counter intuitive if assumed to be over time. One of many examples where the practitioners struggle with the currency of the data specification is:

> *"No, because they can have several current ones but assignments could now only be current assignments and not assignments over time"*

The use of informal language in the scenario, rather than a clear and unambiguous normative language is also a problem. The scenario must be lacking in detail or it is simply a transcription

exercise. On the other hand, removing the detail of the required normative language means that the scenario cannot be interpreted.

Some evidence suggests that practitioners working in a team will respond differently than those working singly since they can check with each other that issues do not make sense. In a laboratory situation, a practitioner working alone would be inclined to attempt to produce the best 'answer' rather than to 'make sense' of the situation because of the laboratory situation. This difference in social organization would also explain why the practitioners spent much longer on the scenario than the previous student subjects. The motivation of the two groups is entirely different, which is one reason why students generally do not act as practitioners.

Most importantly, it is clear that the scenario did not make sense to the practitioners:

> *"I think the root of the problem is potentially that the brief is wrong … The brief is wrong … It must be wrong"*
>
> *"I would say in the real world you really want to hold skills that are not currently being used because otherwise if you don't know people's skills how can you assign them? This is a sort of chicken and egg. People only have skills when they are assigned as using them but how can they be assigned unless they have them?"*
>
> *"The restriction doesn't make sense …I think we're in agreement that some of the restrictions here seem … bizarre in the real world and difficult to model. Well in the real world what you would do in this situation is you would go talk to the business users and you would get to the bottom of this. You would ask them, you wouldn't just read this brief and try to blindly model, you'd say, we really need to understand exactly what you mean by that."*

The researchers who used this scenario made assumptions that do not fit with practice. Creating and selecting a reasonable scenario seems to be a more complex task than assumed. Probably the scenario builders made a positivist assumption that a diagram can mirror a scenario as a representation of the text, whereas the practitioners need to be able to do more. The practitioners must be able to make sense of the scenario, which is an interpretivist issue.

This argument raises questions about how scenario users scored the models developed in laboratory experiments. The score presumably partly reflects the lack of sense that the scenario makes to the readers. Chaiyasut and Shanks [1994] and Shanks [1997] are perhaps the only scenario researchers to try to account for this issue in their research method. In their work, the individual modellers were able to ask questions about the scenario during the experiment.

A final speculative point concerns the reason why this scenario was thought to be valid by the researchers. Could it be the case that ternary thinking led to the creation of a domain that did not make sense?

WAS THE INTERPRETIVE RESEARCH SUCCESSFUL?

Previous researchers imposed various frameworks on their research methods. The lack of a rigorous initial framework in this research method is a potential weakness that means that the findings must be taken in context. However, a direct interpretation successfully showed some rich insights into the use of conceptual modelling. This paper presents the first research to examine how a team of practitioners model a scenario. It produces findings that cast doubt on the

practicality of conceptual entity-relationship modelling with n-ary relationships. These findings taken in context with other work involving argument and practitioner surveys, for example, make a convincing case, particularly in the absence of relevant conflicting data. Obviously a large body of theory weighs against the work of three practitioners. The research method used can only point to the need for more research, rather than providing definitive proof. This limitation is a basic weakness of the interpretive method. The success of the interpretive research is in learning that theory does not yet account for practice.

The interpretive method finds a situation where the team modelling process is one of discussion and positing sentences. These sentences are constrained by a normative language that makes business sense by conforming to the rules established by the field of linguistics. Practitioners use their method to explore and make sense of a scenario and to expose things that do not seem reasonable. The diagram reflects the normative language and specifies the agreed version of a conversation, much like the minutes of a meeting. The discussion is more important than the diagram.

It was initially useful to use the interpretive method with a constrained scenario because it made the interpretation simpler. The disadvantage of the constrained interpretation is that the findings do not extend to business users. Moving the interpretive method further into the practice domain would mean interpreting a set of real modelling sessions with business users. This approach, howevedr, would raise various practical difficulties, particularly the amount of dialogue that would need to be interpreted. However, the questions raised about social construction by business users would only be accessible in a practice context.

Practitioners' expertise in modelling relationships is different from the conceptual modelling that is taught to undergraduates. The implications for teaching are that students can use interpretive research findings to understand how conceptual modelling is used. The practice of modelling can be studied and understood in its own context. It is important for undergraduates to gain a clear idea of how conceptual modelling theory is applied and when it is useful. Students could learn as much from studying transcripts from practice as they can from completing modelling exercises themselves.

VI CONCLUSION

An interpretive case study was used successfully to learn about the practicality of using entity-relationship modelling in a ternary situation. The findings can be compared with previous studies of novice modellers who used the same scenario. Previous research excluded the interaction between practitioners that would occur during a modelling session because novices were used to complete tasks in isolation. A team of experienced practitioners are shown to use entity-relationship modelling in a business context of social interaction about design. The interaction proves to be a key part of the modelling process. Practitioners 'talk with the notation' as well as using the notation to draw a diagram. The entity-relationship model constrains the social interaction because the model provides a way of talking about design. The practitioners use the model to talk about a normalized relational data structure in a way that undermines the idea of the entity-relationship model as an independent conceptual model. The findings show that theories from the field of linguistics explain why the model is used in this design dependent way and suggest that this dependency may be inevitable. When the design conversation is about a normalized relational data structure practitioners do not benefit from using a special notation for ternary relationships. On the contrary, the practitioner's design dependence seems to enable them to expose aspects of a domain that do not make business sense.

The wider implication is that interpretive research is important in generating insights about the extent to which conceptual modelling is usable by practitioners. Interpretive research highlights the importance of being able to distinguish between ideas about conceptual modelling and ideas about how to apply modelling to practice.

Editor's Note: This article was received on September 18, 2002 and was published on March ___, 2003. The paper was with the author for approximately 2 months for four revisions.

REFERENCES

Barker, R. (1989) *Case*Method: Entity Relationship Modelling*. Boston MA: Addison Wesley

Batini, C, Ceri, S and Navathe, S (1992) Conceptual Database Design: An Entity Relationship Approach. Redwood City, CA: ,Benjamin Cummings

Batra, D. and Anthony, S. (1994) Effects of Data Model and Task Characteristics on Designer Performance: A Laboratory Study. *International Journal of Human-Computer Studies* (41) pp.481-508

Batra, D. and Davis, J. (1992) Conceptual Data Modelling In Database Design: Similarities and Differences between Expert and Novice Designers. *International Journal of Man-Machine Studies* (37) pp.83-101 (Also a prior version as: Batra, D and Davis, J. (1989)Conceptual Database Design by Novice and Expert Database Designers. *Proceedings of the Tenth International Conference on Information Systems* pp. 91-99)

Batra, D., Hoffer, J.A. and Bostrom, R.P. (1990)Comparing Representations With Relational And EER Models. *Communications Of The ACM* 33(2) pp. 126-139

Batra, D. and Marakas, G.M (1995) Conceptual Data Modelling In Theory And Practice. *European Journal Of Information Systems* (4) pp.185-193

Carter, P. D., Patrick, J. D., and Deane, F. P. (2001) EXCOVE and Using Videos in Knowledge Elicitation. *International Journal of Human Computer Studies*, (54)3, 301-317.

Chaiasut, P. and Shanks, G. (1994) Conceptual Data Modelling Process: A Study of Novice and Expert Data Modellers in Halpin, T. and Meersman, R. (eds.) *Proceedings 1st International Conference on Object-Role Modelling*, University of Queensland.

Chen, PP-S. (1976) The Entity Relationship Model: Towards a Unified view of Data, *ACM Transactions on Database Systems* 1(1) pp. 9-36

Dey, D., Storey, V.C. and Barron, T.M (1999) Improving Database Design Through The Analysis Of Relationships. *ACM Transactions on Database Systems* (24)4 pp. 453-486

English, L., P. (1999)*Improving Data Warehouse And Business Information Quality.* New York: Wiley

Hitchman, S (1995) Practitioner Perceptions on the Use of Some Semantic Concepts in the Entity-Relationship Model. *European Journal of Information Systems,* (4) 31-40

Hitchman, S. (1997) Using DEKAF To Understand Modelling In The Practitioner Domain. *European Journal of Information Systems*, (6)3 pp.181-189.

Hitchman, S. (1999) Ternary Relationships – To Three or Not to Three, Is There a Question ? *European Journal Of Information Systems* (8) December pp.224-231

Hitchman, S. (2000) Object-Oriented Modelling In Practice: Class Model Perceptions In The ERM Context *Proceedings of the 19th International Conference on Conceptual Modelling, ER2000*, Edited by Liddle, S.W, Mayr, C and Thalheim, B. Salt Lake City, USA 9-12 October 2000 (published as Springer Verlag Lecture Notes In Computer Science Vol 1920, ISBN 30540-41072-4) pp.397-408

Hitchman, S. (2002), The Details Of Conceptual Modelling Notations Are Important – A Comparison Of Relationship Normative Language, *Communications Of The Association Of Information Systems* (9), pp. 167-179

IDEF1X (1993) Federal Information Processing Standards Publication 184, December 21, www.idef.com/Downloads/pdf/Idef1x.pdf

Jackendoff, R (2002) *Foundations Of Language* Oxford: Oxford University Press

Johnson, R., A. (2002) Object-Oriented Systems Development: A Review Of Empirical Research, *Communications Of The Association Of Information Systems* (8), pp. 65-81

Kim, Y-G. and March, S.T. (1995) Comparing Data Modelling Formalisms, *Communication*s of the ACM* (38)6 pp.103-113.

King, N (1995) The Qualitative Research Interview pps.14-36 in Cassell, C and Symon, G (eds) (1995) *Qualitative Methods in Organizational Research: A Practical Guide* London: Sage Publications

Kock, N., Gray, P.,Hoving, R.,Klein, H.,Myers, M. and Rockart, J. (2002) IS Research Relevance Revisited: Subtle Accomplishment, Unfulfilled Promise, Or Serial Hypocrisy? *Communications Of The Association Of Information Systems* (8) pp. 330-346.

Nijssen, G.,M., Duke, D. J. and Twine, S., M. (1990) The Entity-Relationship Data Model Considered Harmful in *Proceedings of The Sixth Symposium on Empirical Foundations of Information Systems and Software Science*. October 19-21, 1988, Atlanta, Georgia., pp.109-130 Published as Zunde, P. and Hocking, D. (eds) *Empirical Foundations of Information and Software Science V*, New York: Plenum Press

Russo, N. L and Wynekoop, J. L. (1997) Studying System Development Methodologies: An Examination of Research Methods. *Journal of Information Systems* **7** pp.47-65

OMG (1999) *The Unified Modelling Language Specification* Version 1.3 June 1999, Needham, MA: Object Management Group http://cgi.omg.org/cgi-bin/doclist.pl

Searle, J. (1995) *The Construction Of Social Reality* London: Penguin

Shanks et. al.(1993) The Role of Experience in Conceptual Data Modelling *Proceedings 4th Australian Conference on Information Systems*, Brisbane: University of Queensland

Shanks, G (1997) Conceptual Data Modelling: An Empirical Study of Expert and Novice Data Modellers. *Australian Journal of Information Systems* (4**)**2 pp. 63-73

nks, G. Darke, P. and Broadbent, M (1998)Successfully Completing Case Study Research:Combining Rigour, Relevance and Pragmatism *Information Systems Journal* (8) pp. 273-289

Shoval, P. and Frumermann, I. (1994) OO and EER Conceptual Schemas: A Comparison of User Comprehension. *Journal of Database Management* (5)4 pp.28-39

Shoval, P. and Frumermann, I. (1997) OO and EER Conceptual Schemas: A Comparison of User Comprehension Concept Modelling *ER'96 15th International Conference on Conceptual Modelling*, Lecture Notes in Computer Science 1157, Berlin: Springer Verlag

Shoval, P. (1997) Experimental Comparisons Of Entity-Relationship And Object-Oriented Data Models. *Australian Journal of Information Systems* (4)2 pp. 74-81

Shoval, P. and Shiran, S. (1997) Entity-relationship and Object-Oriented Data Modelling - An Experimental Comparison of Design Quality. *Data and Knowledge Engineering* (21**)** pp.297-315

Topi, H and Ramesh, V (2002) Human Factors Research On Data Modelling: A Review Of Prior Research, An Extended Framework And Future Research Directions. *Journal Of Database Management* (13)2 3-19

Veres, C. and Hitchman, S (2002)Using Psychology to Understand Conceptual Modelling in Wrycza, S. (ed.) *Proceedings of the Xth European Conference on Information Systems*, Gdansk, Poland, 2002, p. 473-481.

Walsham, G (1995) Interpretive Case Studies in IS Research:Nature and Method. *European Journal of Information Systems* (4) pps. 74-81

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. 495-528.

Wand, Y and Weber, R. (2002) Research Commentary: Information Systems and Conceptual Modelling - A Research Agenda *Information Systems Research*, (13)4363-376.

APPENDIX I. THE SCENARIO PARAGRAPH

The scenario for this transcript is taken from Shoval and Shiran [1997). A similar scenario was used several times by others [e.g. Batra et al., 1990, Shoval and Frumermann, 1994, Shoval and Frumermann, 1997, Shoval 1997). The scenario used is an extract from a larger scenario – only the modelling of this part of the scenario is reported in this paper.

> Many workers can work on a project. A worker can work on many projects, but can be assigned to only one project in a given city. It is necessary to track the date on which a worker began working on a project in a given city. We are interested in the city name and population for each city. A worker can have many skills (e.g. preparing material requisitions, checking drawings etc.), but he/she may only use a given set of skills on a particular project. A worker uses each skill that he/she posses in at least one project. It is necessary to keep track of the number of hours that a worker uses each skill in a project. Each skill is assigned a number. A short description is required to be stored for each skill. Projects are distinguished by project numbers. It is required to store the estimated cost of each project (in $).

APPENDIX II. . PRACTITIONER RESPONDENT'S EXPERIENCE

	Respondents		
	1	2	3
Which ERM notation do you consider yourself experienced in ?	Barker	IBM's SDM, Barker	SSADM, Barker
Which case tools do you consider yourself experienced in ?	Designer	Designer	Designer, ERWIN
How many days of ERM training did you have in higher education ?	2	0	0
How many days of formal practitioner training in ERM ?	6	5	3
… of which how many days were case tool related ?	3	0	3
How many years experience of ERM ?	3	10	2
How many years of ERM as your main work activity ?	2	2	3 months

How many different ERM projects have you worked on ?	15	20+	3
What were the approximate number of entity-types involved in your last three project models ?			
Project 1	428	420	240
Project 2	240	50	420
Project 3	253	40	30
Do you consider yourself reasonably expert at ERM	yes	yes	yes
Do you consider yourself still learning about ERM	yes	yes	yes

APPENDIX III. MODELLING SESSION TRANSCRIPT

Practitioners spent an hour on the first two paragraphs of the complete scenario before they reached this stage of the scenario. They originally read through the whole scenario at the beginning of the exercise. Time is measured from the point at which the practitioners first directly modelled this third paragraph. Actor 1 is generally drawing on the whiteboard. The transcript follows.

Time	Action	Actor	Description of diagram action and transcript extract
0			Start reading and modelling the paragraph
01:03	1		Create project entity
01:17	2		Create city entity, with discussion of the scope of the 'worker' concept – does this mean the worker sub-type or workers in the company generally ? This was left as an unresolved assumption that the relationship was with the worker entity-type
2:12		1	So, many workers can work on a project
2:17	3		Create the project/worker relationship as a line
		2	But a worker can work on many projects
		1	There's a clue in the name (meaning 'assigned')
3:02	4		Remove original worker-project relationship (3) and create the link entity assignment and relationships.
			Interpretation Point 1 *The scenario authors had intended the sentence 'A worker can work on many projects, but can be assigned to only one project in a given city' to flag a ternary relationship between worker, project and city. The practitioners use the previous (first) sentence 'many workers can work on a project', together with 'a worker can work on many projects' to derive a many to many relationship between worker and project before they consider the ternary information. The practitioners use a 'link' entity-type 'assignment' - many-to-many relationships are always modelled this way by these practitioners.*
		3	There's an attribute of date to put on.
3:45	5		Create attribute date in assignment
		2	We are interested in city name and the population of each city
4:04	6		add attributes city name and population
		3	It doesn't say whether a project is wholly within a city

Time	Action	Actor	Description of diagram action and transcript extract
		1	No, it doesn't. You think that … Is name the identifier for a city?
			An exchange of city names that are known to be in several countries.
4:53	7		Add id as identifier for city
5:10			(Discussion of the population for city as historical data – how to deal with changing values per project assignment ? Decided to assume to just retain a current city population)
			Interpretation Point 2 *The practitioners soon spot that although the scenario attempts to provide unique identifiers the city identifier is missing. A key problem then arises about whether this is a 'snapshot' or a 'historical' situation. Are we keeping track of city populations over time ?*
5:24	8		Create relationship city/project (crow's foot on project end)
5:48		3	Working on a project in a given city, you're saying that a project only takes place in a given city and doesn't span cities.
		1	That was the way I read it
		2	It doesn't actually say that
		1	Well, no. What it says, strictly speaking, is that a worker can only work in a city it doesn't say
		2	No, it doesn't say that – a worker can work on many projects but can be assigned to only one project in a given city
		1	Yes, one project in a given city so does that mean that the worker works in the one city or that the project is in the one city ?
		3	No, it means one worker in the city can only be assigned to one project, that project can span many cities
6:30		2	There is something a bit complicated here.
		1	There is a great deal of ambiguity in the text
			Interpretation Point 3 *One problem for the practitioners is that they do not know enough (from the scenario) about projects and cities. Trying to resolve this brings the practitioners back to the ternary sentence at which point they realize that there is a complex constraint issue here. It seems difficult for the practitioners to make sense of the worker, city, project constraint without also being able to make sense of the binary relationships between each of these entity-types. In order to understand 'working on a project in one city' the practitioners need to be able to understand how projects relate to cities. There is evidence throughout the transcript that thinking 'three ways' is only possible (if at all) on the basis of understanding the individual interactions between the pairs of entity-types involved. A ternary relationship notation, on the other hand, hides or precludes these individual interactions*
6:39		3	We could make an assumption here
		1	Shall we make an assumption that it's simple and move on ?
		3	No,
		1	Or make an assumption that there is a many to many relationship between project and city
6:52		2	This model will not enforce the rule that a worker can only work on one project in one city. This will allow the worker to work on many projects and these projects would all be in the same city
		1	Ok, so let's not call this project then …
7:10	9		creation of link entity to reflect many-to-may relationship between project and city, named project location, this now relates to assignment. Involves creating new project entity

Time	Action	Actor	Description of diagram action and transcript extract
			Interpretation Point 4 *The 'project-city' link entity-type, location, will eventually disappear. The practitioners are intuitively following the 'rule of thumb' for checking fourth and possibly fifth normal form issues around entity-types with multipart keys. Here they are posing questions that the scenario inventors did not consider. This is going to be a crucial cause of problems for the practitioners as some of the information they consider important is missing. When the scenario inventors 'simply' used a ternary relationship they forgot about specifying facts considered important by the practitioners concerning the possible binary interactions between the three entity-types involved. The wording of the scenario is problematic in that the practitioners have to imply the situation about 'one project in one city', 'one project in several cities', 'two project in one city' based on 'only one project in a given city'. The normative language of the scenario is not complete with respect to the specifications needed for the diagram.*
		1	This now says that a worker can only be assigned to one of those rows (project location) so a worker can only work in one city that the project is in
		2	No, it doesn't, that still doesn't enforce it
		1	No (agreeing) because there's a many there" (meaning project location may have many assignments)
8:28	10		delete the crow's foot (project location may only have one assignment).
8:32		1	Are these things the same thing ?" (based on one to one relationship)
		2	The worker can only be assigned to one project in one city so should the assignment be to the project
		1	No, isn't it just that these two things that we called assignment and project location, aren't these the same thing ?
			Interpretation Point 5 *An assignment to one project location is a way of constraining the worker to one project city and this would seem to enforce the required constraint. The practitioners are, however, incorrect to constrain project location to one assignment – there should be many assignments, or workers at a project location. Before this is realized the practitioners focus on the one-to-one relationship that they just created. The practitioners know that a one to one relationship will very probably result in the entity-types location and assignment merging into one entity-type. Alternatively a one-to-one relationship will be incorrect and flag some flaw in the thinking. The discussion that follows is in the context of both of those possibilities. It is important to realize that at this point the practitioners are less concerned with whether the diagram is correctly specifying what they think they understand, but is instead acting as a way of talking through different possibilities. Deciding immediately whether the one-to-one relationship is reasonable is less important, here, than talking around the issues raised by it. The one-to-one relationship is posited to enable the issue of working in one city to be explored from a different angle.*
		2	We're back to where we were before
		1	No we're not. Are we ?
9:18		1	What we are saying is that there can be multiple occurrences here (assignment) so a project can be worked on in these cities and what we are saying is that a person can only be assigned to one of those
		2	the worker can only be assigned to a city once
		1,3	no
		2	can be assigned to many cities but can only be assigned to one project in the city .. so if we could make the association between worker and city we .. could enforce the one city only rule, it is then getting the association to the project
10.25		3	I think if we go with what we've got though, can't we just make city part of the key of assignment

Time	Action	Actor	Description of diagram action and transcript extract
		2	yes, is it to do with the uniqueness of the key in assignment, so if we make city part of the primary key, yes, of assignment then that would ensure it would only appear once.
11.06	11		Add relationship city assignment
			Interpretation Point 6 *By adding the relationship to assignment the practitioners are still positing a specification – they are going to see what happens now before considering the one-to-one relationship. With the benefit of looking ahead it is clear that Actor 2 is correct in identifying that the constraint required is concerned with the key in the entity-type assignment but also wrong to assume that including the city in the assignment key will enforce the required constraint. A worker could still be assigned to a city on several projects. Whilst it might be best to avoid incorrect assumptions it would constrain the modeling session considerably (at this stage) to be too careful. The idea here is to think around the issue and work through different ideas until a point is reached that makes it worthwhile to re-track to check that the diagram now makes sense. Bear in mind that the existing one-to-one relationship is also know to be suspect so it may not be worth re-tracing until this disappears. So during a modelling session discussion the diagram may contain several untested assumptions that are used to think around the domain. The exploration of the domain may be much more important than the final diagram.*
			A discussion then follows based on the new triangular relationship indicating an error in the model. 1 suggests removing relationship between city and project location
		2	well let's see if that fails to satisfy something else in the brief
11:54	12		remove relationship project / city
		2	do we need the project location now ?
		1	I'm not sure we do now, this is really looking like project again, isn't it ?
11:50	13		remove project and re-name project location to project *Interpretation Point 7* *The practitioners have not included the cardinality of the 'one' end in the diagram at this point. Although the diagram shows a solid line (indicating a mandatory 'one' end) the 'one' relationships are all optional – this is just an aversion to drawing dashed lines on the whiteboard. Two of the 'crow's feet' are also missing from the diagram – there are many assignments for both cities and projects. It is not that the practitioners have forgotten this but rather that they are still thinking through the constraint issue – this is not a finished specification. The is one of the points in the session that the model reaches a more stable state and the practitioners will recheck their understanding of this version.* *The practitioners used the one-to-one relationship (incorrectly defined) in order to merge two entity-types and have now specified an assignment entity-type that is directly equivalent to the ternary relationship in the scenario builder's answer. The practitioners are able to think about the idea of an assignment (an event compared to, say, a worker) and talk about the assignment in the same conceptual way as they talk about all of the entity-types - assignment becomes a part of the conversation. The use of assignment in the conversation is the strongest evidence for a ternary notation being a redundant concept. The question here is not about how well the concept fits 'reality' - but about how useful 'assignment' is when making sense of the scenario. It may be possible to theorise about some ontology where assignment is a 'different sort of thing' – a ternary relationship – compared to, say, worker. It would be harder to decide whether an assignment was a different sort of thing than a project. However, this is of no relevance if worker, project and assignment are all usefully used as conversation concepts.*

Time	Action	Actor	Description of diagram action and transcript extract
			The issue of the 'one city' constraint has not yet been solved, although Actor 2 does later provide a specification for this. To enforce the constraint a unique key on just city and worker is required – although this prevents any chance of a worker being later reassigned back to the city. One of the practitioners does provide a correct constraint solution later in the session. Alternative unique keys are not shown on the diagram notation but can be documented in the case tool. This is an interesting case of issues that are considered so unusual that it is better to annotate the diagram or make a note in the documentation rather than raise a new notation. Having lots of notations for unusual circumstances would hinder the modeling process but give a little used return. The practitioners eventually conclude that the 'one' constraint is unreasonable. It seems that the scenario inventors have created an artificial situation in order to provide an example for using a 'one' notation. When such notations are put into scenarios it not surprisingly proves difficult to make business sense of the scenario.
		1	So this is now saying that there is an intersect between workers and projects, so many workers are assigned to projects and a project can have many workers, each of those project assignments is in a city. Therefore you could deduce all the cities a particular project is worked upon
		3	You've only recorded what's happened, that's fine, (but) if you want to record what's going to happen …
13:10	14		Add crow's feet and 'relationship is part of primary key' indicator (I) to assignment. Also add #ID, estimated cost and (name) to project. (Note – discussion that each project would have a name even if not in scenario).
			Interpretation Point 8 *This is the point where the practitioners have re-tracked and completed the assignment entity-type that almost represents the ternary relationship that the scenario specifies. The practitioners have done a lot of investigation around the binary relationships that they have used to further their understanding of the domain. The use of the 'I' notation on a relationship specifies that the unique identifier from the 'parent' entity-type is included as a unique identifier in the 'child' entity-type. This kind of constraint is not discussed in the academic literature and has never been used in published scenario solutions. The practitioners are using some tacit assumptions about what this 'I' means when they interpret the domain and the 'I' notation soon has an important impact on the model later. The practitioners are still having problems making sense of the lack of information about 'snapshot' versus 'over time'.*
		3	A worker can have many skills ...(reading from scenario)
14:58	15		Add skill, worker skill
		3	A worker can have many skills but he/she may only use a given set of skills on a given project (seems to be reading from notes, but not a quote)
		2	There's a relationship between the assignment and the worker skill
		1	Indeed (re-draws diagram to make room for relationship)
		1	A worker has a number of skills, does that mean a worker used a skill set on a particular project assignment ? Or is it just a number of their skills are used on a project assignment .. that's suggesting ..(draws implication)
17:09	16		draws relationship from assignment to worker skill

Time	Action	Actor	Description of diagram action and transcript extract
			Interpretation Point 9 *The practitioners have now begun to deal with what was intended as the second ternary constraint. The practitioners use the 'I' notation in a way not considered by the scenario inventors. The practitioners decide to relate the initial (worker, project, city) ternary entity-type instead of directly linking the new entity-type with the 'original' entity-types of project and worker. This is because the practitioners know that worker skill will inherit the keys from assignment – or to put it another way this is a 'lower level' relationship than was intended by the scenario authors.* *The practitioner interpretation is that worker-skill is for this assignment - it is a 'list' of worker skills on this assignment and not a list of the worker skills actually owned by the worker. Worker skill is therefore a misnomer – assignment skills would be a better name. With worker-skill we can track skills per project, per worker and per city. This interpretation satisfies the scenario but adds some extra information to the model. This is an example of the way that practitioners will pro-actively add to the 'meaning' of the modeled domain. This is equivalent to the scenario authors having linked skill to the worker, city and project ternary relationship.* *A potential problem is that we will have to decide on skills per project-city and not on skills per project, i.e. a worker may be programming on a project in Dublin and programming and testing on the same project in Paris. This is not unreasonable since we ought to know who was doing what and where, but it is more detailed that the scenario specified.*
		1	.. (continues) that this question marks that
	17		Draws question mark on relationship worker / worker skill
		1	.. (continues) relationship then. If you don't have that relationship what we are saying is that workers don't have skills unless we are using them on a project, but there is a line in the brief that says ..(interrupted)
		2	Yes, a worker used each skill that he / she possesses in at least one project
		1	Yes, so that's saying that we don't need that relationship (worker / worker skill)
		2	No, I think you do – you need to know the total, ah, this is interesting, I can see what you are saying, but the assumption would be that someone somewhere knows whether all the skills of an individual are being used or not, but how do you know what the total skills of an individual are ?
18:04		3	If you are a new employee with seven skills on just one project only using two skills how do you know what the others are ?
		1	That's not what the brief says. The brief says a worker uses each skill that he/she possesses on at least one project. So that's saying that all the skills that a person has are currently being used.
		2	I think that's probably what they're saying
		1	Let's get rid of that relationship. I would say in the real world you really want to hold skills that are not currently being used because otherwise if you don't know what skills people have how can you assign them ? This is a sort of chicken and egg. People only have skills when they are assigned as using them but how can they be assigned unless they have them ?
		2	Yes
		1	But following the brief I think we are saying that relationship is redundant according to the brief.
19:14	18		remove relationship worker / worker skill
19:22	19		skill no attribute added
		3	We missed number of hours
19:52	20		Skills description added

Time	Action	Actor	Description of diagram action and transcript extract
	21		Add I on worker skill relationships
			Interpretation Point 10 *The practitioners are starting to understand why the scenario doesn't make sense to them. It is self evident to the practitioners that 'definitional data', as they later call it, would be needed in order to do anything useful with the data. To make sense of the domain it is not reasonable that we do not know what skills a worker has, and yet we can assign a worker to a project with particular skills – how would we know what these skills were ?. It becomes clear to the practitioners that the suggestion in the scenario that skills are known about 'per project assignment' is unreasonable. The scenario builders have used a 'simple' ternary relationship which has resulted in no thought being given to workers and skills – this is also the issue of 'testing for fourth normal form' that doesn't get done when a ternary solution is proposed. Use of a ternary relationship will hide the analysis of this issue. The diagram at action 21 is close to the final version of the diagram. After the final action (26) the diagram is the same but the 'I' on the project-assignment relationship is removed and the attribute 'time spent' is added to the entity-type worker skill. It will take half as long again to establish these small changes.*
19:59		2	It is necessary to keep track of the number of hours that each worker uses each skill on a project
		1	That complicates this (assignment / worker relationship). This simple relationship doesn't support that, we have another entity.
		3	Something between worker skill and assignment
20:20		3	Well, you've posted your assignment down to worker skill
		2	Yes, but if you held it at worker skill you'd only be able to … it doesn't say that the worker only uses their skill once
		3	No, but you haven't …
		1	No, I'm going with him (3) on this 3 has suggested that time just goes in there (worker skill)
		2	But that would just be the total time for that skill across all assignments
		3,1	No, it would not
		1	No, this (worker skill) inherits the key of assignment (pointing to diagram)
		2	Oh, that's true
		1	So, this thing here (worker skill) each one of these is one skill being used on one project
		2	Yes, of course, ok
		1	So, the time spent is there (worker skill)
21:12	22		Add time spent attribute to worker skill
		1	And this says that this relationship (worker skill / worker) wasn't only redundant, you cannot have that relationship if you've got time spent there
		2	Yes, that's true
21:30		1	Right, shall we just independently read through the whole thing and see if there is anything we haven't covered

Time	Action	Actor	Description of diagram action and transcript extract
			Interpretation Point 11 *Again, adding an attribute to the diagram results in some validation. At this point Actor 1 is attempting closure on the modelling session – but the other practitioners are going to bring the discussion around again to the ternary constraints. There are still another 12 minutes of iterative discussion that covers old ground and 'history' or 'over time' and the lack of definitional data are the main causes of the iterative discussion. One of the practitioners re-reads the 'mission statement' at the beginning of the brief. In one sense the discussion becomes 'bogged down' – however, another interpretation would be that the modelling technique enforces discussion of various alternatives until a satisfactory resolution is found and the scenario makes sense. Here the discussion is wide ranging because of the scenario shortfalls. The lack of domain experts to question is obviously critical since this could otherwise have resolved the outstanding questions.*
		3	Well, there are things we haven't covered. He or she may only use a given set of skills on a particular project
		2	Sorry, what's the point you are trying to make ?
		3	It's kind of implying that a project has a particular set of skills that's required and a worker may only use a given set of skills on the project.
		2	We've got that using (points to worker skill)
		1	Nah
		3	No, you're recording what has happened by this mechanism, not what …
		1	Yes
		2	Yes, but in terms of the brief I think it's covered, in the real world you wouldn't want to model it that way
		3	Well, if you read the first line again "an engineering firm, requires a database to keep track of all employees; their skills, projects and departments" – it doesn't say that we've done you might want to keep track of another project that is due next week, you might have a project next week. So you might assign some workers to it.
22:54		1	There is that particular line in the brief that says he or she may only use a given set of skills on a particular project and that's saying the skills that were used (pointing to worker skill) on the project there's nothing definitional here that they're the only ones that could be used but the fact is that since we are saying that skills are not related to workers and that they only exist at the point of assignment we can't have a structure in there that says you have another skill that …
		2,3	No, no
		1	Oh, you can. You can because what we are saying there is that potentially, over time at least, I'm not sure, is a person only assigned for one project at once
		2	No, they can work on many projects but only once within a city
23:40		1	Right, so we could have another worker here (worker skill) who has a skill that say accountancy, for example, supposing they are an accountant and an architect, they could be an accountant on one project an architect on another (pointing to assignment) and in there (worker skill) we would have one item of this would be them being an accountant on project one and another one with them being an architect on project two. What stops them being an architect on project one
		2	The unique key of assignment

Time	Action	Actor	Description of diagram action and transcript extract
			Interpretation Point 12 *The practitioners are using examples of instances to follow through their discussion and examples are often documented in the case tool by these practitioners. The entity-types are perhaps easier to think about when they are instanced. Although examples do not appear on the final diagram they will be written next to entity-types from time to time on the whiteboard. It is also worth noting that the practitioners mix all three elements of the high level framework levels proposed by Batra & Davis (1992) in the same sentences. This may reflect the difference between a team session and a talk-aloud protocol in the laboratory. Rather than a movement between the representation, recognition and enterprise levels suggested by Batra & Davis (1992), the practitioners within the team environment seem to mix whatever helps make the sentences useful.*
		1	Yes, but what we're saying is, or what 3's implying is, there should be something in here (the whole diagram) that says these are the only skills that this worker is allowed to use on this project. There's an implication in the brief that we need some definitional data that says which particular skills are allowed to be used on a project
24:36		3	I mean time spent doesn't really fit very well there, does it?
		2	No (agreeing)
		3	Time spent should be in the assignment
		1	No, because that would way how much time a person had spent working on a project, this is particularly saying ..
		3	No, on the assignment
		1	..(continuing) they want a break down, they want to know how much time a person has spent on each assignment but using each skill. If you may have two skills ..
			Interpretation Point 13 *Actor 1 here is beginning to talk about the business processes that will use the data specified by the model to decide whether it is reasonable to have the fine grained association to assignment rather than to project. The cross checking of data model and process model would form a key task of model checking and refinement outside of the scenario situation. It is also interesting that the practitioners are using the placement of an attribute to make decisions about the model entity-type structures. The next part of the transcript is a detailed discussion of attribute placement, instigated by Actor 3. Small changes in the model, swapping a crow's foot for example, can imply a lot of changes to the domain semantics. Practitioners gain a lot of tacit business domain meaning from the notation that may not be apparent to a non expert – a small change in the notation can have a large effect on meaning.*
		3	yes, but I'm talking about the relationship as well
		1	(gives whiteboard pen to 3) I'm not following quite what you are saying
25:20	23	3	(gets up to whiteboard) I think time spent should live there (assignment) and I think that this (moves crows foot on assignment / worker skill, writes time spent on assignment and leaves original time spent in worker skill). We want that sort of structure there
		2	Yes, but that would be time spent on all their skills on a given assignment
		3	No, you've got a skill, this intersect (assignment) says you've got this worker with this skill in this city on this project and has spent this amount of time
		2	Oh, you've turned that one around (crow's foot on assignment / worker skill). No, I don't like that. The problem with doing that is that you loose … again, you're allowing the person to work on different projects in the same city using different skills
26:10		3	Yes, but
		2	And That's not what the brief says
	24	3	but it's this time thing isn't it. That's got to go here (assignment) that's got to be part of the key. Flags the # on time as part of key

Time	Action	Actor	Description of diagram action and transcript extract
		2	yes, but if you make the time part of the primary key that allows you to assign the worker to multiple projects in the same city
26:26		1	3's right as what 3's saying then is that you are inheriting worker skill as part of the identifier of this assignment and you could then add ..
		3	Does it say that
		1	Yes, That's right 3 it's an obscure issue
		2	The other issue I've got is the direction of that relationship (project / assignment) I think there is something wrong here. So we need to look at that in a bit more detail as well
		3	(Reads from brief) "A worker can work on many projects, but can be assigned to only one project in a given city" It should say at any one time
		1	Even so
		3	That's ok if we make time part of the key
		2	No, because we haven't got end date of the assignment so you need to make sure ..
		3	It's implied by this (assignment)
		2	We're assuming we don't need to keep history as well
		1	I think you need to put that (relationship assignment / worker skill) back around the way it was before. I don't support that change
27:29	25		Put assignment / worker skill relationship back as it was and remove time spent as an attribute from assignment (attribute still in worker skill)
			Interpretation Point 14 *The next part of the transcript covers the point when Actor 2 realises that they have still not specified the initial constraint to only work on one project in one city. Actor 2 suggest that to make sense of the scenario a worker really just works in a particular city once. There is no point in worrying about what project they are on – the constraint is simpler in that a worker can work in a particular city just once. In a sense this is an attempt to get to the business rationale for only working in one city – what is the point of this constraint ? If a worker is on two projects wouldn't it make more sense for them to be in the same city if possible ? Without an understanding of the rationale for this constraint the practitioners cannot make sense of it. This is not an objective transcription of the scenario to a diagram – it is an attempt to interpret why the constraint is there. This is a reflection of a fundamental difference between business analysis and what a student might be motivated to do in a laboratory situation..*
		2	OK, no, the problem I've got with .. I think we need to sort out this other one first (project / assignment)
		1	I'm missing what you think is wrong
		2	Well, the problem I've got with this is that we've said that you can only work on one project in a given city, yes ?
		1	Yes
		2	Although you can work on multiple projects now I think what you have to do, Oh, I see, project becomes part of the key doesn't it. No, that's no good, that would allow you to work on two projects in one city. We want to enforce the rule that they can only work on one project so you have to make project a foreign key into assignment (means just a foreign key, not part of the primary key)
		3	Yes
		1	What about the I (the included in primary key symbol on project / assignment) – so you're saying the primary key is only employee and city ?
29:50	26	2	Yes. So project becomes optional (ie not in the primary key). Remove I from assignment / project relationship

Time	Action	Actor	Description of diagram action and transcript extract
		3	So you're saying a worker can't work in the same city twice, ever
		2	No, I'm saying at a given .. Whilst they are working on one project in a given city they can't work on another. Now, it is quite reasonable to change the project .. no you'd create a new assignment.
			Interpretation Point 15 *Actor 2 eventually specified the correct constraint. Actor 1 next attempts to overrule the new constraint based on the tacit knowledge that the primary key of an entity-type like assignment is always a three part key. The scenario is implying that only current data will be held to reflect the current (snapshot) view of assignments. For example an employee would be located on different projects in the same city twice over time even if they are allowed to be there only once currently. So the solution to the constraint problem founders against the unspecified issue of currency.*
		1	No, unfortunately,2, you've not done this right. You were quite right before that with project as part of the key you couldn't enforce that a worker could only work in one city on one project at once. But now you've removed that as part of the primary key. 3 is quite correct and what you've now enforced is that a worker that can only ever work in one city once, irrespective of the project because now to put them on another project in the same city would give you a duplicate assignment
		2	Yes, so we might need an assignment history
		1	So, you're now saying that assignment is current assignment
31:00		2	Yes. No, because they can have several current ones but assignments could now only be current assignments and not assignments over time
		3	And they've got to be in different cities
		2	That's right
		1	Well, you would think actually that someone in the real world could only actually work in one city at once. They could be able to work on many projects rather than be able to work on many cities but only on one project
31:27		2	Yes
		1	So, it does seem like the brief there is potentially the wrong way round, because what you would think this was showing well if someone is working in Birmingham then they can't possibly be working on a project in Edinburgh just because the physical location ..
		2	I think we've probably found the root of our problem here
		1	I think the root of the problem is potentially that the brief is wrong
		2	The brief is wrong
		1	It must be wrong
			Interpretation Point 16 *Here the practitioners decide that their understanding of the scenario does not match their experience or their view of what would be reasonable in the 'real world'. Their modelling method has not allowed them to quickly build an 'elegant' diagram with complex notation. Instead their method has acted as a device to assess the reasonableness of the domain description. To put this another way – it is more important to see if the model makes sense than it is to model what the scenario says. 'Making sense is clearly an important aspect of modeling, indeed it may be the key aspect of the method – and suggests that measuring diagram 'correctness' against model answers in experiments has partly been measuring the perceived unreasonableness of the scenario. The time spent by the practitioners in 'making sense' of the scenario may explain why they took longer than the previous student research subjects. There is a now last discursive check about abandoning the modeling session.*
		3	No, you could be working in one city on an email system but be acting as a consultant to another

Time	Action	Actor	Description of diagram action and transcript extract
			project
		2	What are the sort of skills
		1	But, if that was true. Yes, I see what you mean, 3. What you are really saying is that city is nothing to do with the project at all. City is something to do with the current location of the employee. And what we are really saying is that an employee can only be located in that city
32:28		2	You're making an assumption that someone is physically located on that city, though, aren't you. There's nothing here to say that people are going to move from city to city
		1	OK, but if someone is not physically located in a city, if this is not what this is all about, why would you allow people to work on multiple projects based in multiple cities but not allow someone to work on two projects in the same city
		2	I agree, but we don't know what company we are dealing with
		1	The restriction doesn't make sense
		2	Well, it doesn't say in the sort of headline introduction it says, the engineering firm needs a database to keep track of all employees, their skills projects and departments. It says nothing about employee location
		1	You're quite right. I'm just trying to explore what would be a real restriction. Because I think we're in agreement that some of the restrictions here seem
		2	Very bizarre
33:20		1	… bizarre in the real world and difficult to model. Well in the real world what you would do in this situation is you would go talk to the business users and you would get to the bottom of this. You would ask them, you wouldn't just read this brief and try to blindly model, you'd say, we really need to understand exactly what you mean by that.
33:52		2	Yes
			Interpretation Point 17 *The practitioners have now reached the stage where they believe the scenario doesn't make sense – so it would not be worthwhile trying to complete the diagram without more information. We are left with the interesting situation that the practitioners lacked an 'elegant' way of dealing with the scenario – a ternary relationship notation – but used their method to uncover a domain that didn't make sense. If we compare this to the 'invented answer' which failed to establish 'reasonableness' with the ternary notation, we could conclude that the notation used by the inventors hindered them from realizing that the scenario didn't make sense. We would also conclude that, given these are experienced practitioners, the occurrence of this kind of constraint, particularly involving a 'one' constraint (one city) is rare at best as it seems to be outside their experience. Although the modelling method used lacks a clear notation for dealing with the 'one' constraint this is exactly the part of the model that seems to make no sense anyway. The suggestion is that such a notation is as invented as the need to have it. We would further argue that there is no point is creating an extra notation for even a rare situation since this would make the notation harder to use generally when some simple annotation (or alternative unique key, for example) would suffice.*

APPENDIX IV. POWERPOINT PRESENTATION

It is possible for the reader of the transcript in Appendix III to re-create the diagram (labeled Figure A-1 on the next page) from the transcript. Furthermore, all of the diagramming actions can be replayed in the Powerpoint slide presentation that can be found at http://cais.isworld.org/articles/11-26/model%20build3.pps. Once the first chart appears, press the PgDn and PgUp keys to go forward and back. As shown in the figure, the last action is (26).

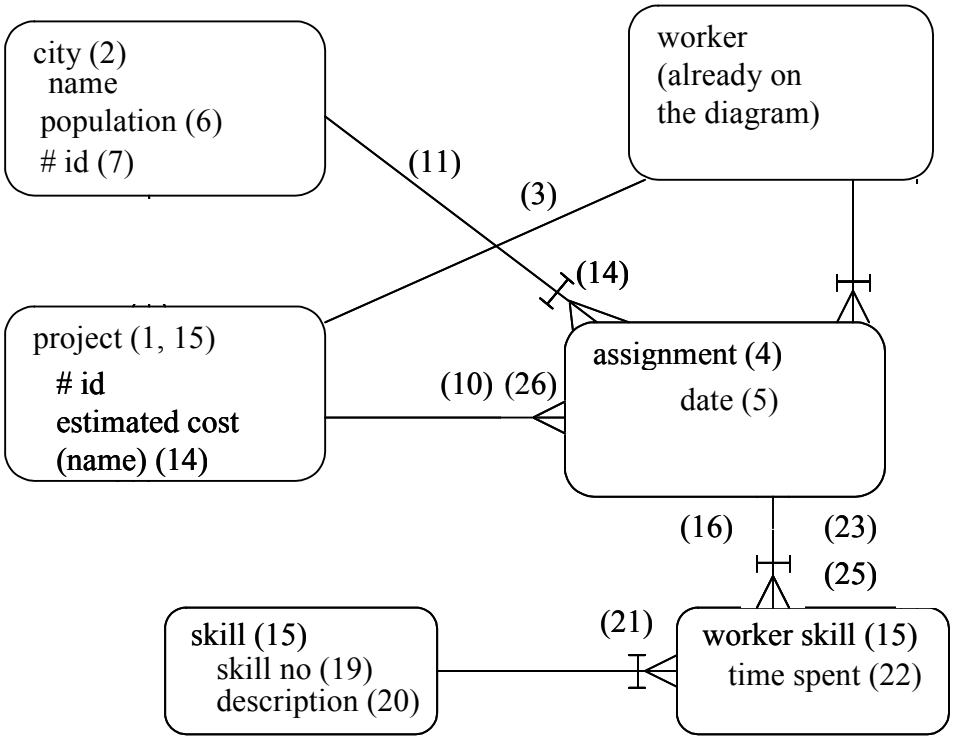


Figure A-1. Diagram Discussed in the Transcript

Copyright © 2003 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from ais@gsu.edu

ISSN: 1529-3181

EDITOR-IN-CHIEF
Paul Gray
Claremont Graduate University

AIS SENIOR EDITORIAL BOARD

Cynthia Beath Vice President Publications University of Texas at Austin	Paul Gray Editor, CAIS Claremont Graduate University	Sirkka Jarvenpaa Editor, JAIS University of Texas at Austin
Edward A. Stohr Editor-at-Large Stevens Inst. of Technology	Blake Ives Editor, Electronic Publications University of Houston	Reagan Ramsower Editor, ISWorld Net Baylor University

CAIS ADVISORY BOARD

Gordon Davis University of Minnesota	Ken Kraemer Univ. of California at Irvine	Richard Mason Southern Methodist University
Jay Nunamaker University of Arizona	Henk Sol Delft University	Ralph Sprague University of Hawaii

CAIS SENIOR EDITORS

Steve Alter U. of San Francisco	Chris Holland Manchester Business School, UK	Jaak Jurison Fordham University	Jerry Luftman Stevens Institute of Technology

CAIS EDITORIAL BOARD

Tung Bui University of Hawaii	H. Michael Chung California State Univ.	Candace Deans University of Richmond	Donna Dufner U.of Nebraska -Omaha
Omar El Sawy University of Southern California	Ali Farhoomand The University of Hong Kong, China	Jane Fedorowicz Bentley College	Brent Gallupe Queens University, Canada
Robert L. Glass Computing Trends	Sy Goodman Georgia Institute of Technology	Joze Gricar University of Maribor Slovenia	Ruth Guthrie California State Univ.
Juhani Iivari University of Oulu Finland	Munir Mandviwalla Temple University	M.Lynne Markus Bentley College	Don McCubbrey University of Denver
Michael Myers University of Auckland, New Zealand	Seev Neumann Tel Aviv University, Israel	Hung Kook Park Sangmyung University, Korea	Dan Power University of Northern Iowa
Nicolau Reinhardt University of Sao Paulo, Brazil	Maung Sein Agder University College, Norway	Carol Saunders University of Central Florida	Peter Seddon University of Melbourne Australia
Doug Vogel City University of Hong Kong, China	Hugh Watson University of Georgia	Rolf Wigand University of Arkansas	Peter Wolcott University of Nebraska-Omaha

ADMINISTRATIVE PERSONNEL

Eph McLean AIS, Executive Director Georgia State University	Samantha Spears Subscriptions Manager Georgia State University	Reagan Ramsower Publisher, CAIS Baylor University