

Do Viewpoints Lead to Better Conceptual Models? An Exploratory Case Study

Steve Easterbrook, Eric Yu, Jorge Aranda, Yuntian Fan,
Jennifer Horkoff, Marcel Leica, and Rifat Abdul Qadir
Department of Computer Science
University of Toronto, Toronto, Canada
bul-khp@cs.toronto.edu

Abstract

The use of viewpoints has long been proposed as a technique to structure evolving requirements models. In theory, viewpoints should provide better stakeholder traceability, and the ability to discover important requirements by comparing viewpoints. However, this theory has never been tested empirically. This paper reports on an exploratory case study of a key hypothesis of the viewpoints theory, namely that by creating separate viewpoint models to represent different stakeholder contributions, and explicitly merging them, important hidden requirements can be discovered. The case study compared two modelling teams using the i^ notation to capture requirements for new web-based counselling services for a large charitable organisation. One team used viewpoints; the other did not. The conclusions include that viewpoint merging improves the understanding of the problem domain, but is very time consuming. The process of merging was more important than the merged product. The study also indicates a need for better model management tools, as both teams encountered difficulty in managing large, evolving models.*

1. Introduction

In viewpoints-based modelling, participants are able to maintain their own (partial) models of the system and its requirements, without being constrained by other participants' models [5]. By keeping the viewpoints of different stakeholders separate, analysts can identify and explore the relationships between them, and participants can understand one another's perspectives better [6, 2].

A key feature of viewpoints-based approaches is toleration of inconsistency [4]. While many agent- and goal-based modelling languages allow different stakeholders' needs to be separated in the model, they assume the model should be consistent overall. In contrast, viewpoints allow analysts to build and modify many partial, overlapping models, without maintaining consistency between them. In

principle, this idea can be applied to any conceptual modelling language.

The underlying theory of viewpoints is this: When approaching a conceptual modeling problem, it is better to build many fragmentary models representing different perspectives than to attempt to construct a single coherent model. The theory suggests that viewpoints should bring the following benefits:

- Stakeholder buy-in and traceability. By capturing separately different stakeholder viewpoints during elicitation, stakeholders can identify their contributions, and requirements can be traced back to their source.
- Structuring the process. Viewpoints permit parallel development of separate 'workpieces', with no constraint on consistency between them, so the modelling process can be distributed amongst a team of analysts.
- Delayed commitment. Viewpoints allow alternative representations of the problem, so analysts can delay choices about which aspects are important, and how they should be modelled, until the stakeholder's perspectives are better understood.

A corollary of the theory is that conceptual disagreements are best handled by an explicit process of comparing viewpoints. This is not self-evidently true. For example, a competing theory, based on research on negotiation, suggests that if participants identify too closely with the positions they start out with, they can become too entrenched for constructive negotiation [7]. An alternative to viewpoints modelling, then, is to concentrate the modelling activities on the areas of consensus between perspectives, and resolve differences informally, as the models are first constructed.

The use of viewpoints brings new challenges, such as how to identify relationships between viewpoints, and how to discover and handle inconsistencies. Therefore, the benefits claimed for viewpoints-based modeling have to be weighed against the extra cost of managing inconsistency. However, we are aware of no empirical studies that investigate the basic tenets of the theory of viewpoints, nor the

scope of its applicability. The theory remains untested. To address this gap, we conducted an exploratory evaluation of some of the key hypotheses of the theory of viewpoints.

The case study we describe in this paper concerns the modeling of a large, charitable organisation, Kids Help Phone, using the *i** modelling language. The study compares two modelling teams, only one of which used viewpoints to structure its models. To allow for detailed comparisons, both teams worked from the same set of stakeholder interview transcripts. Thus, the study concentrated on the modelling activity itself, and ignored potential uses of viewpoints during the initial information gathering.

Our aim was to explore what differences the use of viewpoints make to conceptual modelling. We set out to explore the hypothesis that explicit comparison of stakeholder viewpoints would yield a richer understanding of the problem situation, including the discovery of hidden assumptions and requirements. We also wanted to understand other aspects of the theory. For example, how would the viewpoints-based models differ from other models? How hard is it to compare and merge viewpoints? What additional needs are there for tools to support the viewpoints-based approaches?

2. Viewpoints

Viewpoints have been discussed in the Requirements Engineering literature for at least twenty years. Unfortunately, different authors have used the term ‘viewpoint’ for widely different things. Viewpoints have been used to mean entities in a system’s environment [10], different classes of users [18], to distinguish between stakeholder terminologies [21], and to partition the requirements process into loosely coupled workpieces [15]. Darke & Shanks [1] provide a survey and comparison.

An emergent theme is that ‘viewpoints’ provide a technique for partitioning a large quantity of information collected from many different sources. The information is collected in coherent, but overlapping chunks (‘viewpoints’). Because the concepts included in different viewpoints can overlap, there is the potential for inconsistency [20]. However, inconsistencies between viewpoints can be dealt with separately from the task of describing and elaborating each viewpoint. This toleration of inconsistency distinguishes viewpoints from other problem structuring techniques.

Most of the early work on viewpoints emphasized the benefit they offer during *elicitation*. Viewpoints can be identified with stakeholders, with classes of users, with individual analysts, and so on, to address the multiple perspectives problem [6]. Each viewpoint owner is then free to describe her contribution using whatever notation and problem decomposition she chooses, and to focus on the aspects that matter most to her.

A number of frameworks have been proposed that provide explicit support for identifying, tracking and resolv-

ing inconsistencies between viewpoints [4, 8, 17]. There is no consensus on when inconsistency should be eliminated. For example, van Lamsweerde *et al.* [22] concentrate on resolving inconsistency at a very early stage by resolving divergences between stakeholder goals, while Nuseibeh *et al.* [14] argue that some inconsistencies are never resolved, even in an operational system.

More recent work has focussed on the problem of managing inconsistency between viewpoints. A number of representation schemes have been proposed for capturing and managing the consistency relationships in modeling languages. These include a first order logic for checking XML documents [13], a production rule approach for checking UML models [12] and a structural mapping technique based on graph morphisms for graphical notations [19]. Other work has explored formal reasoning techniques that tolerate inconsistency [9, 3].

3. Study context

Kids Help Phone (KHP) is a non-profit social service organization that provides counselling to kids and parents across Canada through the phone and the web. KHP is actively seeking to adapt its services to take advantage of technology advances and to respond to changing societal needs. In particular, many counsellors believe that kids are increasingly likely to prefer the internet to the phone when seeking advice and help, and KHP needs to adapt its services to remain relevant to these kids.

For non-profit social service organizations such as Kids Help Phone, the challenge of introducing new internet services is even greater than in the business world. There are fewer established practices to draw from. The success of KHP relies on the cooperation and goodwill among many participants, including volunteers, professional counselors, executives and donors. The target users are diverse, differing in age group, family situation, issues faced, language and culture, youth subcultures, communication and technology skills. A wide range of technologies are possible, with various modes and degrees of interactivity. Finally, it is difficult to evaluate success, due to the overriding need to protect the anonymity of the clientele.

Senior management at KHP approached researchers at the University of Toronto in the fall of 2003, seeking advice on developing new internet-based services. We proposed to use the *i** framework [25, 24] for a systematic analysis of the organizational setting and the requirements for strategic technology change. The project was funded by Bell Canada (one of the major donors to KHP) through its Bell University Labs program.

*i** was appropriate for this problem because it emphasizes the analysis of strategic relationships among organizational actors. *i** involves two types of model: a *Strategic Dependency (SD)* model in which actors are related by de-

pendency links to other actors, and a *Strategic Rationale (SR)* model, which elaborates the SD by exposing the reasoning within each actor, identifying goals, tasks, resources, softgoals, and beliefs (generically known as *intentional elements*), and their relationships (means-ends, task decompositions, softgoal contributions). The graph constitutes a network by which goal achievement can be evaluated by a label propagation procedure.

We selected this project as an ideal case study to explore the viewpoints theory for a number of reasons. First, the project would clearly involve a diverse set of stakeholders, with competing goals, and so naturally fits within the scope of the theory. Second, the management of KHP were highly motivated to participate in the proposed requirements analysis, so we anticipated a high degree of access to key stakeholders. Third, the proposed requirements modelling was of the entire organisation, thus providing a natural way to scope the case study, and allowing us to avoid complications that may arise when trying to isolate a business activity from some larger corporate context. Finally, the project would represent one of the largest applications of the *i** modelling language to date, and we considered that the use of viewpoints might help address some of the anticipated scalability challenges.

4. Methodology

4.1. Why a Case Study?

Case studies are an important empirical method, suitable for investigating questions that cannot be addressed through controlled experiments. Whereas controlled experiments rely on statistical analysis over a large number of instances, case studies rely more on qualitative analysis to connect cause and effect. They are particularly suited to studies in which the researcher has little control over the key variables.

The key hypotheses underlying the use of viewpoints in requirements modelling cannot be tested experimentally, because of the difficulty in controlling the variables from one treatment to the next. Essentially, the benefits of viewpoint modelling are only likely to be evident for large scale modelling, under conditions that cannot be replicated in the laboratory. In particular, the study of viewpoints cannot be separated from the organizational context in which they are used, and the effects may take weeks or months to appear.

We used an *exploratory case study* as the basis for our research design [23]. Exploratory case studies are ideal for analyzing what is common and/or different across cases that share some key criteria. They are appropriate for preliminary studies in which it is not yet clear which phenomena are important, or how to measure these phenomena.

In our case, we were particularly interested in understanding how the use of viewpoints would affect the mod-

elling process. While the theory of viewpoints suggests some specific benefits, these have not yet been observed empirically. Not enough is known about how exactly viewpoints are best deployed, nor how the expected benefits arise. For these reasons, it would be premature to try to measure the cost/benefit trade-off. For this study, our intention was to explore *how* viewpoints affect the modelling process.

4.2. Hypotheses

Although exploratory case studies do not necessarily begin with specific hypotheses, we did derive several hypotheses from the theory of viewpoints, to guide the study design.

Our central hypothesis was: “*Modelling stakeholder viewpoints separately and then combining them leads to a richer understanding of the domain*” We took “richer understanding” to mean that we would see evidence of hidden assumptions, disagreements between stakeholders, and potential requirements revealed through the use of viewpoints, which otherwise would go unnoticed.

Additional hypotheses were as follows:

- *Viewpoints modelling improves traceability to individual stakeholders.* Because the comparison and merging of viewpoint models is carried out explicitly, it should be easier to see how the models were derived.
- *Viewpoints modelling improves readability of resulting models.* By readability, we mean the ability of the original stakeholders to comprehend the models – the hypothesis arises from the observation that viewpoint models should remain faithful to the stakeholder’s own conceptualization of the problem, and this should carry through even when viewpoints are merged.
- *Viewpoints modelling improves the ability to capture divergent and minority opinions.* This hypothesis arises because without viewpoints, we would expect to see conservatism during the modelling process – a modeller will tend to ignore information that does not fit the model she is developing.
- *Viewpoints modelling makes team modelling easier because it decomposes the modelling task.* We interpret this to mean that the decomposition into partial, overlapping viewpoints offers additional advantages over any partitioning and projection techniques available in the modelling language.

4.3. Study Design

Our central hypothesis predicts a difference in the level of understanding of a problem situation achieved through the use of viewpoints. To investigate this, we needed to compare modellers using viewpoints with those not using viewpoints, for the *same problem domain*. As we had some control over the requirements modelling activities, we took

the opportunity to set up a comparative study, with two separate teams developing their own models of the problem. Each team's modelling activities constituted an embedded unit of analysis within the case study.

As a starting point for the modelling, we interviewed the key stakeholders from across the KHP organisation. All interviews were conducted by the same pair of project members, to ensure consistency of interview style. The interviews were structured around a basic set of questions, but the interviewees were encouraged to raise any other topics they felt relevant. Each interview lasted an hour, and each was recorded¹ and transcribed. Interviewees were subsequently shown the transcripts of their interviews and invited to make comments and corrections.

We conducted 14 stakeholder interviews, covering all major roles in the organisation, including CEO, senior management, counsellors, operational managers, information technology specialists, human resource management, and fundraising. For practical reasons we were unable to interview any potential users of KHP services, but we were able to interview two student ambassadors, who help to promote awareness of KHP in their schools and local communities.

The transcripts from these interviews (approx. 140 pages in total) were used as a baseline dataset by our two modelling teams. Each team consisted of three modellers, with varied experience of the i^* notation. All team members were graduate students at the University of Toronto, conducting thesis research in requirements engineering. They were all aware of the intent of the study, and participated in identifying the hypotheses and the study design itself.

To control for modelling expertise and familiarity with the domain, we ensured that each team contained members with previous experience of i^* , and each team contained one of the interviewers. In addition, all project members participated in an initial modeling exercise, based on the Montreux Jazz Festival case study described in [16].

The two teams were as follows:

1. The global modeling team (G team for short). This team was instructed to develop a single large i^* model of KHP, using all the transcripts. All members of the team worked together on the model², allocating modelling subtasks between them as appropriate.
2. The viewpoint modeling team (V team for short). This team was instructed to develop individual models (viewpoints) of each stakeholder interviewed, and then to merge them to obtain a model of the entire organisation. Each team member took a share of

1 One interviewee declined to be recorded.

2 One team member left the project shortly after the study started, leaving us with only two members in this team, but we do not believe this affected the balance of modelling expertise and domain familiarity.

the transcripts, and developed initial viewpoint models without conferring. The viewpoint merging was conducted by the whole team working together.

To minimize the impact of this study on the KHP organisation, we asked the modellers to work exclusively from the transcripts, and delayed stakeholder validation of the models until after the teams had completed this stage of the study. In fact, the models turned out to be so large that extensive stakeholder validation was not feasible. Instead, we conducted a workshop with KHP management, in which we presented interesting findings from both teams' models, and invited discussion of the findings. We did not differentiate between the two teams during this workshop, as the goal was to move our analysis of KHP requirements forward, rather than to further the aims of this case study.

Several observations can be made about this study design. First, although the design shares some features of a controlled experiment, we did not regard it as such, as we did not have the resources to replicate each treatment. Further, there were additional confounding variables that we did not attempt to control. For example, we did not control the order in which the V team members tackled their viewpoint models; one would expect some modelling bias to be introduced by having the same team member build viewpoint models of several different stakeholders. Similarly, we did not control the experimenter bias that arises from having the modelling teams being active researchers on the project.

Second, the study design limits the potential benefits of viewpoints to just the *construction* of conceptual models based on the interview data. It would be interesting to extend the viewpoints idea back into the interview process itself, so that each modeller interviewed just those stakeholders whose viewpoints she was to model. However, such a design would prevent the comparative study of the G and V cases. Similarly, we would have liked to extend the viewpoints idea forward into at least one validation cycle with the stakeholders. We ruled this out for similar reasons.

The application of viewpoints to i^* also raises some interesting issues. The i^* ontology is intentional – it deals with actors' intentions, and how these might be achieved, often through other actors, and with the help of different technologies. Viewpoints may disagree on goals, on means to ends, and on relationships among them. Variations in vocabulary may belie deeper differences in perspectives and values. These issues are more challenging than inconsistencies among different versions of ER diagrams or statecharts. Intentional models are harder to elicit and validate, and tend to be more subjective, as stakeholders are literally disclosing what is at stake for them.

4.4. Data Collection

Our goal was to discover the ways in which the viewpoints modelling differs from the global modelling. Hence

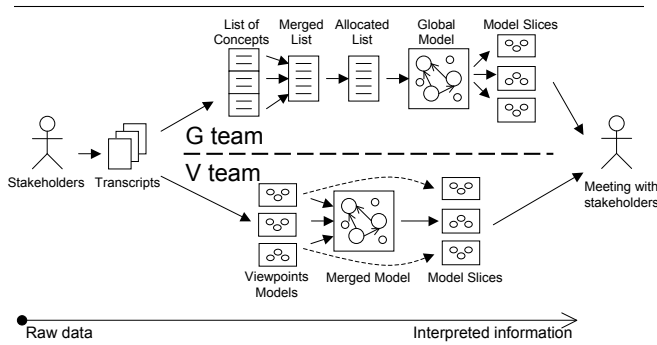


Figure 1. Comparing the processes of the V and G teams

the data collected was purely qualitative, and relied extensively on the subjective notes of the participants. Participants were asked to keep careful notes of what they did at each step of the modelling process, and to record any problems encountered, as well as their reflections on the quality of their models. In addition, we video-recorded some of the viewpoint merging sessions conducted by the V team.

The main comparison between the teams’ models was conducted at the end of the modelling stage, when each team presented its models to the entire research team. At this stage, we looked for concepts present in one team’s model(s) but not the other’s. We also looked for differences in how concepts were modelled, and asked the teams to explain how these elements of the models were developed.

5. Results

In this section, we first provide an overview of each teams’ modelling activities, and then compare the two. Figure 1 illustrates the two modelling processes, indicating the intermediate artifacts produced by each team. The horizontal axis on this diagram indicates relative distance from the raw data.

5.1. G Team Modelling Activities

The G team began by reading all 14 transcripts, highlighting potential i^* model elements. Different highlighting colors were used for different types of i^* element, including: goals, softgoals, tasks, actors, resources, beliefs, and contribution links. At this stage, they focussed more on the actors and intentional elements, rather than the links.

They then extracted all the highlighted text from the transcripts, and placed them in lists categorized by element types. They pruned these lists to remove irrelevant items, and to merge elements that were similar, for example where different stakeholders had mentioned the same concept. They then allocated each intentional element to one or

more of the potential actors. The result of this exercise was a list of approximately 950 intentional elements, and around 120 potential actors and roles.

It was clear that drawing a single strategic rationale (SR) diagram containing all these elements would be impractical. The team therefore divided up the model into a number of separate views. Each view focussed on one or more actors, showing the strategic rationale for these actors, but collapsing the remaining actors to show only strategic dependencies. These views represented projections of a single large SR model, although this was never constructed explicitly. This process resulted in a list of 9 separate SR views.

The team constructed each view by inserting the intentional elements from the lists, and then adding links. Most links added at this stage were not mentioned explicitly in the interview transcripts, but were judged to be sensible assumptions. Some additional intentional elements were added, where it made sense to help connect other elements.

Finally, they reviewed the lists of intentional elements, to ensure that all elements were accounted for. This step entailed more “merging” of elements, when an element was already represented in the model by a similar element. The team also cross-checked the views to ensure they were consistent with one another, and generated a strategic dependency (SD) model from each of the SR views, and finally a single SD model for the entire organisation.

5.2. V Team Modelling Activities

The V team divided the 14 transcripts between the three team members (4 to 5 transcripts each). Where there were similar stakeholder roles and positions, these were evenly distributed among the three members; otherwise the distribution was random. Each team member created a model for each stakeholder transcript assigned to him/her. They did not discuss their models at this stage, although they did share questions and answers to general modeling issues.

The team used two principles to help ensure each viewpoint was faithful to the original stakeholder: (1) each model should only contain information present in the transcript, so that information remembered from other transcripts is excluded, and (2) the models used the same vocabulary as the stakeholder, or a close paraphrase.

They then attempted to merge the resulting viewpoint models, over a series of team meetings. Some of these were conducted by two team members, while the third was away. They started with the most important issue for KHP’s service planning: the counselling role itself. The dataset contained interviews with three different counsellors, so each team member had a counsellor viewpoint model.

In each merging session, the team started by selecting an element in the model that seemed to be shared by all viewpoints. For example, in one case, they started with a high-level softgoal, and reviewed all contributors to that softgoal.

In another case, they started by reviewing all dependencies among two agents (counselors and kids) in each viewpoint.

The team constructed a merged model by including elements that matched across the viewpoints, together with any elements only mentioned in one viewpoint. If the elements differed in level of detail, the most detailed version was used. Where the same term had been used for different concepts, one of the terms was changed to make the difference in meaning clear.

Often, a concept appeared in two or more viewpoints, but was expressed in different ways. For example, one model had “web services” as an agent, another had them as an aspect scattered among several agents. This difference in structure made merging them very difficult. To make things worse, such problems tended to appear in basic definitions. For example, different counsellors gave different definitions of “counselling”. In such cases the team attempted to merge pieces that were not conflicting, and just flag the conflict in the rest. Sometimes, they developed a new structure, if they decided none of the original viewpoints adequately captured the shared view.

During the process, the team used the original transcripts and the context of each element in the viewpoint models to help decide whether concepts matched. This exploration was time-consuming. Sometimes the team reached agreement on the best way to merge mismatching elements, other times they did not. In that case, they merely noted that the conflict could not be solved based only on the raw data.

5.3. General Observations

Both teams found it hard to extract model elements from the text. Often ideas were described in many sentences, which they had to summarise or paraphrase, possibly losing meaning or misinterpreting the idea. This was exacerbated by our study design, as it did not permit further discussion with stakeholders over how to interpret their interview comments. On the other hand, this may be normal for any organisation where access to stakeholders is limited.

For larger models, the only practical method of viewing and editing them was to divide them into a number of separate *views*, each stored in a separate model file. These views overlapped, but were intended to be consistent with one another. Hence, they are not *viewpoints*, according to the theory. Nevertheless, the effort of ensuring consistency between them was considerable, and was done manually, due to lack of tool support. A change to one view often meant that many other views also had to be updated.

Even these views were too complex to use directly with the stakeholders in our subsequent workshops. Instead, we extracted *slices* from the views. A slice is a well-defined subset of a model, extracted using a slicing algorithm; we developed several such algorithms as a result of this project [11]. A slice is selected to address a specific analysis

question. However, this required difficult judgments about which analysis questions were most pertinent.

We used Microsoft Visio for the modelling. The tool was reliable and handled even our largest models gracefully. However Visio does not handle some *i** syntax well, especially linking intentional elements. We would have saved time if a robust, syntax-aware editor was available for *i**.

5.4. Comparing the G and V modelling

We will now contrast the experiences of the two teams.

Model size was the main challenge for the G team, given the volume of data of the combined set of interviews:

- The lists of intentional elements extracted from the transcripts were too long to be manageable. It was difficult to check for similar items.
- It was difficult to decide how to divide the large models into cohesive views.
- To obtain workable models, they had to split some large actors into smaller roles. For example, the counsellor actor was split into the roles: Provide counselling; Counselling training; Counselling Information provider; etc. Such splits often seemed artificial, forced by modelling practicalities, rather than real problem domain concerns, so it was hard to decide to which role each intentional element should belong.
- The size of the models led to serious layout problems. The models became so cluttered that it was hard to add new elements without re-arranging and/or restructuring them.
- Once the models grew in size, they became hard to read, so for example it was hard to tell if a given element or link had already been added.
- It was hard to review and validate the models because viewing them was so difficult. Even printed on 36" wide paper, the largest views were still hard to read.
- It was hard to analyze the models using the *i** goal evaluation method, again because they were so large.

The V team avoided most of these problems, because the viewpoint models were significantly smaller than the global model, and usually smaller than the G team’s views. The viewpoint models were easier to build initially, and easier to view and edit subsequently.

Table 1 compares the model sizes. This comparison is intended to indicate the scalability problem only. Model size does *not* indicate relative completeness (conceptual models are always incomplete), nor the number of requirements identified (modeling in this study was used to help understand the domain, before any requirements scoping was attempted). It is not meaningful to sum the sizes of the V team viewpoints, because they overlap considerably.

<i>i</i> *element	G team		V team
	big SR model	view average	viewpoint average
Actors	61	13	10
Goal + Softgoals	943	118	60
Tasks	530	66	47
Resources	57	9	20
Goal Contribution links	1013	113	50
Means-Ends links	150	16	15
Decomposition links	303	34	46
Dependency links	437	72	28

Table 1. Sizes of the models

Backwards Traceability was much easier for the V team. They could rapidly identify a point in a transcript when defending a modelling choice in their models, either from memory or via a simple keyword search. This was clearly facilitated by their careful use of the stakeholders' own vocabulary.

The G team had difficulty doing this, and report that they probably only ever sought to trace back to the transcripts about 5 times in the entire modelling process. The G team models were somewhat distanced from the original transcripts. For example, there is a softgoal "Support Individuality in Counselling Techniques", but the term "individuality" does not occur in the transcripts. G team's lists of model elements indicated which interview this goal came from, but not where within the transcript. Eventually, we traced it to the following passage:

"...Each person counsels from a very particular place and that's their tool. ... When you start telling people what they should have done you are interfering with their ability to be themselves ... All it does is make people on the phone who sound like they're reading things from a script...."

The G team's use of lists of model elements acted as an intermediate representation between the transcripts and the models. These lists reduced traceability, but brought other advantages. For example, they served as an early indicator of the size and complexity of various parts of the problem domain, and led to some early consideration that model decomposition would be important. Once model elements in the list were attributed to agents and roles, this indicated agents that were too large and needed to be split. Hence, the lists played an important role in choosing initial decompositions of the modelling problem.

Merging the viewpoints was the main problem for the V team, and they were unable to produce an integrated model of the entire problem domain. The challenges they encountered in merging included:

- Differences in level of detail. For example, one model referred to "information" about a caller, while another broke it down to "age", "gender", "province" etc.
- Differences in modelling style. For example, one modeller represented many ideas as beliefs, another focussed more on high-level goals, and the third focussed on operationalizations and lower-level goals.
- Differences in level of familiarity with *i**.
- Modelling freedom. Often there is a choice whether to represent a concept as a task or a goal. The same problem occurs with agents/actors/positions.
- Differences in vocabulary between viewpoints. Most such differences were introduced by stakeholders, rather than the modellers. Some of these were irrelevant, but others represented a difference of perspective.
- Differences in the scope of each interview.

The team had to make sense of each of these differences during the merge process, frequently referring to interview transcripts, which slowed down the process. The net result was that viewpoint merging was painstakingly slow. However, it did offer greater insights into some crucial aspects of the problem domain, which we discuss in the next section. Also, it seemed to be important to have all the modellers present when trying to merge viewpoints. In the sessions where one member was absent, progress was hampered by many unanswered questions about the missing team member's viewpoint model.

Model Analysis was another difference between the teams. In preparing for the subsequent discussions with the stakeholders, we used the two teams' models in different ways. For example, we applied the goal evaluation procedure to (slices of) the G team's model, to determine which high level goals were satisfied by alternative counselling methods. We did not use goal evaluation on the viewpoint models, because each viewpoint captured only a fraction of the relevant information. On the other hand, the V team viewpoints were very useful for understanding differences of opinion between stakeholders, and some of the issues that surfaced during our attempted viewpoint merges became the focus of our subsequent discussions with KHP.

5.5. Discussion

The problem domain in this case study was sufficiently large that serious scalability issues arose for *i**modelling. The central problem was how to structure large models to make them manageable to view, edit, validate, and analyze.

The use of viewpoints to divide the problem domain by stakeholder allowed the V team to avoid most of the scalability problems. The G team created views to manage the

size of their models. Both types of structuring involve overlap between partial models, and hence redundancy. However, the nature of this redundancy is very different in the two cases. The views created by the G team were designed to fit together, with the relationships between them clearly understood at the outset. If they became inconsistent, this always indicated a mistake that needed fixing. The V team's viewpoints were constructed without any regard for how they could eventually fit together. When they were inconsistent, this revealed interesting differences between the original stakeholders.

Some insight into the effect of viewpoints can be seen by considering at what point modelling commitments are made, and how the teams combined information from multiple stakeholders. The G team resolved many differences of vocabulary when merging their lists of candidate intentional elements. However, the lists were relatively informal, allowing more flexibility, and hence less commitment. By going straight to models, the V team had to make modelling commitments earlier, and then had to resolve these commitments when merging viewpoints. In effect, the V team has to do "sense making" twice, but have the advantage of better traceability to the original data.

In this sense, the benefit of viewpoint merging lies not in the merged result, but in the insights gained in the process. The V team did not finish merging their viewpoints, but this did not matter. In the process, they made several important observations about the problem domain, which we followed up with the stakeholders in subsequent sessions. Several examples illustrate this point:

- The resource "Context Information" appeared in two viewpoints, but had a different meaning in each. In the first, it referred to basic information from callers (age, province, gender,...). In the second, it referred to cues available to the counsellor from his/her interaction with the caller (noise, tone of voice, the way kids refer to themselves, etc.). Exploring this difference yielded a very important observation about how counsellors gather contextual cues in order to give proper counselling. Such cues are vital, and careful thought will be needed for how they will be affected by a move from telephone counselling to other modes. None of the original models adequately conveyed this point.
- One viewpoint presented the softgoals "Anonymity" and "Confidentiality" of the caller as depending upon counsellors. Another viewpoint presented them as depending upon KHP as an organization. Again, the difference sparked a discussion. Our conclusion was that the anonymity and confidentiality that counsellors (as individuals) provide to kids is different from that provided by KHP as an organization. Each has its own particularities that should be considered.

- One viewpoint implied that giving resource information (such as phone numbers of local services) to kids is part of what it means to give counselling; another viewpoint separated these. A discussion started on this topic, since if we can separate the act of counselling from the act of giving resource information, as the first viewpoint suggested, then a website with resources and referrals might be effective. If, on the other hand, when the counsellor gives information he/she is also attempting to discover the caller's real problem, then these concepts should not be separated and a counsellor is necessary. We confirmed the importance of this in a subsequent workshop with counsellors: often a helpful counselling session starts from a call simply intended to get information.

Although there were elements representing these ideas in the G models, their importance was not obvious. We conclude that merging of independently generated stakeholder viewpoints did yield insights that were not available to the G team. However, the process was extremely time consuming, and tended to over-analyze modelling choices.

An obvious suggestion is to use viewpoint merging only very selectively. Unfortunately, our study did not offer any good criteria for where to use it, nor are we able to say whether, in general, the benefits outweigh the costs.

Another suggestion is to reconsider how to approach viewpoint merging. The V team approached it in an "item-by-item" fashion. However, this kind of merging may be too low level. The key insights generated in the V team's merge process are at a higher level intellectually, and come from a reconciliation of concepts, rather than a merging of the model contents. In this case study, these higher level insights arose as a result of the attempt to do low-level merging. It is entirely possible that the same insights can be obtained more readily by comparing viewpoints based on a "big picture" understanding, without resorting to item-by-item merging.

6. Threats to Validity

6.1. Construct Validity

Intentional Validity: Do the constructs we chose capture what we intended to study?

The key construct is idea of a viewpoint. We took "freedom to be inconsistent" as the defining characteristic, which excludes the views used by the G team to decompose their models. However, much of the work on viewpoint consistency management describes exactly this type of model decomposition: a large model is split into fragments that can be edited independently, with tool support handling the maintenance of consistency. We believe that the difference between *views* that are intended to be consistent, and *view-*

points that are not, is a crucial distinction, and the viewpoints research community needs to make it more carefully.

Representation Validity: Do the constructs we chose translate well into observable phenomena?

In this study, we were only able to observe a particular type of viewpoint, associated with interview subjects. Other ways of structuring the problem domain into potentially inconsistent viewpoints are also possible within the theory - for example, by issue or aspect, by organisational role, or even by modeller. Associating viewpoints with individual stakeholders is the approach most mentioned in the literature, so we do not feel this is a serious limitation.

Another important construct was backwards traceability. In this study, we were only able to observe the modellers themselves tracing back to the interview transcripts. Many other participants (developers, testers, users, ...) may want to perform such tracing, possibly to other source artefacts. For these reasons, our conclusions about traceability are very limited, and should be interpreted carefully.

The hardest construct to translate into observable phenomena was “deeper understanding of the problem domain”. We looked for examples of issues raised by each team that were valued by the stakeholders in our followup workshops. The examples cited in section 5.5 may indicate just a difference of emphasis, rather than a deeper understanding. Our best measure is the subjective opinion of the study participants, and the sense of revelation some of us felt when each of these issues was first identified.

6.2. Internal Validity

An important question is whether the differences we observed really are due to the use of viewpoints. Other possible explanations include differences in the participants, the experimenters, the selection process, the procedural setup, the instrumentation, or just coincidental events.

The most likely confounding variable is a difference between the participants assigned to each team. We minimized this by ensuring each team had a mix of experienced modellers, and each team had one of the original interviewers. We did notice problems arising from different levels of familiarity with i^* , but these showed up as differences within each team, rather than differences between the teams.

A major problem with our study design is that the participants were also researchers on the project. This compounds the problem of experimenter bias, because the participants may manipulate the study to obtain the expected outcome. This threat was mitigated in two ways. First, by using an exploratory case study (rather than an explanatory or causal study), we were able to concentrate more on reporting our experience, rather than trying to prove our hypotheses. Further, the only member of the project team who has performed research on viewpoints in the past did not partici-

pate in either of the two modelling teams. Neither of these steps removes this threat entirely; only replication with neutral participants can address this issue.

Finally, it is possible that the instructions given to the participants, and our observations of their modelling activities introduced bias. For example, the constraints placed on either team in terms of structuring their models may have forced an unnatural way of working, hampering their ability to use the full power of i^* modelling. Also, both teams participated in the model review process at the end of the study, and this may have affected how they subsequently reported their own modelling experiences.

6.3. External Validity

The results of this study might not generalize to other modelling projects. In common with much empirical research in software engineering, the participants of our study were graduate students. One could argue that these students represent some of the most knowledgeable experts in the requirements modelling techniques we are investigating. Nevertheless, professional analysts working in a commercial environment may have very different modelling processes.

Also, the results might not generalize beyond the i^* modelling technique. We chose i^* because it is emerging as a leading approach to early requirements modelling, and we were particularly interested in whether i^* would benefit from the addition of viewpoints. Other modelling languages have view structuring built in, derived from their meta-models. This may avoid some of the problems associated with the ad hoc view creation used by the G team. Further case studies are needed to examine the role of viewpoints for such notations.

We also note that the case study began with interview transcripts that are inherently viewpoint-based. Other elicitation techniques may not lend themselves so well to viewpoint modelling. For example, if we started with group elicitation sessions (e.g. JAD workshops), it is not clear that stakeholder viewpoint modelling would even be possible. Group elicitation sessions tend to establish a consensus quicker; it is not clear whether they would also suffer from the problems observed by our G team.

6.4. Reliability

We believe we would obtain similar results if we repeated the study, with one major proviso. The differences experienced by the two teams may be very sensitive to the size of problem domain. Our initial dataset was of such a size that modelling it in its entirety was impractical, whereas modelling each transcript separately was unproblematic. If the individual transcripts were larger, or the overall dataset smaller, some of the differences may disappear.

7. Conclusions

This case study was set up to investigate the role of viewpoints in conceptual modelling. We found that the process of comparing and merging stakeholder viewpoints led to a deeper understanding of the problem domain, and improved backwards traceability to interview data. The viewpoints modelling team also found it easier to cope with overall size of the problem domain.

However, we found no evidence to support our other hypotheses. There was no evidence that the viewpoints models were more readable than the views created by the G team. There was also no evidence that the viewpoints models helped to identify divergent and minority opinions; both teams reported the same differences of opinion amongst the interviewees. However, because the viewpoints models explicitly captured these differences, the V team were able to investigate them more fully, which directly contributed to the deeper understanding reported above.

One surprising finding was that the *process* of merging viewpoints was far more important than the *products* of that merging. This suggests that fully automated merging of stakeholder views is unlikely to be useful. On the other hand, tool support for the process of comparing viewpoints and keeping track of relationships would have greatly facilitated the process of comparing and merging viewpoints.

We also conclude that many of the benefits of viewpoints are contingent upon the nature of the problem situation and the type of analysis desired. In this case study, the viewpoints approach was a good fit, because the stakeholders largely agreed on their high level goals, were extremely committed to the organisation, and placed a higher priority on careful analysis than on early delivery. It is possible that in a problem situation where there is more conflict, or more pressure for a quick solution, other approaches might be more appropriate. The choice of whether to use viewpoints seems to depend on whether we want a deep exploration of the issues, or a rapid convergence to consensus.

Our research on this project continues. We are developing techniques for slicing large models, and better tool support for view management and viewpoint integration. We are also continuing our work with Kids Help Phone, using our models to analyse design choices for their new services. We also plan to seek further case studies to continue our study of the theory of viewpoints.

Acknowledgements: *We thank all the management and staff at Kids Help Phone for allowing us to conduct this case study, and especially to Chris Simmons-Physick for setting up the interviews. We also thank Mehrdad Sabetzadeh for careful comments on an earlier draft. Funding for this work was provided by Bell University Labs (BUL) and NSERC.*

References

- [1] P. Darke and G. Shanks. Stakeholder viewpoints in requirements definition: A framework for understanding viewpoint development approaches. *Requirements Engineering*, 1(2):88–105, 1996.
- [2] S. Easterbrook. Handling conflicts between domain descriptions with computer-supported negotiation. *Knowledge Acquisition*, 3:255–289, 1991.
- [3] S. Easterbrook and M. Chechik. “A Framework for Multi-Valued Reasoning over Inconsistent Viewpoints”. In *ACM/IEEE Int. Conf. on Software Engineering*, pages 411–420, Toronto, May 2001.
- [4] S. Easterbrook and B. Nuseibeh. Using viewpoints for inconsistency management. *IEE Software Engineering J.*, 11(1):31–43, 1996.
- [5] S. M. Easterbrook and B. A. Nuseibeh. Managing inconsistencies in an evolving specification. In *2nd IEEE Symp. on Requirements Engineering*, pages 48–55, York, UK, 1995.
- [6] A. Finkelstein, D. Gabbay, A. Hunter, J. Kramer, and B. Nuseibeh. Inconsistency handling in multi-perspective specification. *IEEE Trans. on Software Engineering*, 20(8):569–578, 1994.
- [7] R. Fisher, W. Ury, and B. Patton. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin Books, 1991.
- [8] J. Grundy, J. Hosking, and W. B. Mugridge. Inconsistency management for multiple-view software development environments. *IEEE Trans. on Software Engineering*, 24(11):960–981, 1998.
- [9] A. Hunter and B. Nuseibeh. “Managing Inconsistent Specifications: Reasoning, Analysis and Action”. *ACM Trans. on Software Engineering and Methodology*, 7(4):335–367, Oct 1998.
- [10] G. Kotonya and I. Sommerville. Viewpoints for requirements definition. *IEE Software Engineering J.*, 7:375–387, 1992.
- [11] M. Leica. Using model slicing techniques to address i* analysis scalability. Master’s thesis, Univ. of Toronto, 2005.
- [12] W. Liu, S. M. Easterbrook, and J. Mylopoulos. Rule-based detection of inconsistency in uml models. In *Wkshp on Consistency Problems in UML-Based Software Development, 5th Int. Conf. on the Unified Modeling Language*, 2002.
- [13] C. Nentwich, W. Emmerich, A. Finkelstein, and E. Ellmer. Flexible consistency checking. *ACM Trans. on Software Engineering and Methodology*, 12(1):28–63, 2003.
- [14] B. Nuseibeh, S. Easterbrook, and A. Russo. Leveraging inconsistency in software development. *IEEE Computer*, 33(4):24–29, 2000.
- [15] B. Nuseibeh, J. Kramer, and A. Finkelstein. A framework for expressing the relationships between multiple views in requirements specification. *IEEE Trans. on Software Engineering*, 20(10):760–773, 1994.
- [16] A. Osterwalder. *The Business Model Ontology: A Proposition in a Design Science Approach*. PhD thesis, Univ. de Lausanne, 2004.
- [17] W. Robinson and S. Pawlowski. Managing requirements inconsistency with development goal monitors. *IEEE Trans. on Software Engineering*, 25(6):816–835, 1999.
- [18] D. Ross. Applications and extensions of sadt. *IEEE Computer*, 18:25–34, 1985.
- [19] M. Sabetzadeh and S. M. Easterbrook. Analysis of inconsistency in graph-based viewpoints: A category-theoretic approach. In *18th IEEE Int. Conf. on Automated Software Engineering*, 2003.
- [20] G. Spanoudakis, A. Finkelstein, and D. Till. Overlaps in requirements engineering. *Automated Software Engineering*, 6(2):171–198, 1999.
- [21] R. Stamper. Social norms in requirements analysis: an outline of measur. In M. Jirotko and J. Goguen, editors, *Requirements Engineering: Social and Technical Issues*, pages 107–139. Academic Press, 1994.
- [22] A. van Lamsweerde, R. Darimont, and E. Letier. Managing conflicts in goal-driven requirements engineering. *IEEE Trans. on Software Engineering*, 24(11):908–926, 1998.
- [23] R. K. Yin. *Case Study Research: Design and Methods*. Sage, 2002.
- [24] E. Yu. *Modeling Strategic Actor Relationships for Process Reengineering*. PhD thesis, Univ. of Toronto, 1994.
- [25] E. Yu. Towards modelling and reasoning support for early-phase requirements engineering. In *3rd IEEE Int. Symp. on Requirements Engineering (RE’97)*, pages 226–235, 1997.