Evaluating Goal Achievement in Enterprise Modeling – An Interactive Procedure and Experiences

Jennifer Horkoff[1] and Eric Yu[2]

[1] Department of Computer Science
[2] Faculty of Information,
University of Toronto, Canada
jenhork@cs.utoronto.ca, yu@ischool.utoronto.ca

Abstract. Goal- and agent-oriented models have emerged as a way to capture stakeholder and organizational goals in a complex enterprise. The complexity of such models leads to a need for systematic procedures to enable users to evaluate and compare the alternative actions and solutions expressed in models. Many existing approaches focus on automated procedures, limiting the ability of the user to intervene. Here, we introduce a qualitative, interactive evaluation procedure for goal- and agent-oriented models, allowing the modeler to supplement the evaluation with domain knowledge not captured in the model. We provide a sample methodology to guide model creation and domain exploration which includes the evaluation of alternatives. We illustrate the procedure and methodology with the i* Framework. Case study experience shows that the procedure facilitates analysis, prompts iteration over model development, promotes elicitation, and increases domain understanding. We describe the results of an exploratory experiment designed to test these findings.

1 Introduction

Goal- and agent-oriented modeling frameworks, such as i* [1], have been introduced in order to model and explore socio-technical domains including actors or stakeholders, their goals and responsibilities, dependencies and alternatives. Although this approach has typically been used as a first step in a system development process, as part of "Early" Requirements Engineering, it can be used more generally as a tool in modeling and understanding an enterprise, including its internal operations and relationships to the external environment. Such models can be used to explore alternative courses of action, analyze their impacts on stakeholders, assess whether stakeholder objectives are met, and can help make tradeoffs among competing goals.

Consider a not-for-profit organization that provides phone counseling for youth, but is interested in reaching more youth using the Internet. Online counseling could be viewed by multiple individuals, and may provide a comforting distance which would encourage youth to ask for help. However, in providing counseling online, counselors lose cues involved in personal contact, such as body language or tone. Furthermore, there are concerns with confidentiality, protection from predators, public scrutiny over advice, and liability over misinterpreted guidance. How can such an organization explore and evaluate options for online counseling?

A. Persson and J. Stirna (Eds.): PoEM 2009, LNBIP 39, pp. 145–160, 2009.

Goal- and agent-oriented models which capture such socio-technical situations often form a complex web of relationships, with alternatives in the model contributing positively or negatively to certain goals, which themselves contribute to other goals. It is useful to assess the level of achievement of a goal in the model when a particular alternative is selected by considering the positive or negative evidence a goal has received via relationships with other goals. However, it can be difficult to trace the effect of a particular alternative on the satisfaction of one or more goals in the model when models contain multiple, multi-step paths of relationships represented with links in the model. There is a need for systematic analysis procedures which consider the effects of alternatives throughout the goal network, providing a consistent way to assign goal achievement levels via propagation along the links.

Models developed to consider enterprises at the goal level are often informal and incomplete, focusing on "soft" goals, such as privacy, which are difficult to precisely define. Such models are intended to be used as sketches, interactive recordings of an ongoing discovery process involving stakeholders and analysts. As the stakeholders express their viewpoints, as discussions occur, and as analysts learn more about the domain, such models undergo continuous change. An ideal analysis procedure would facilitate this process, prompting further discussion or elicitation and iteration over models.

A number of analysis procedures for analyzing goal models have been introduced (for example [2], [3], [4], and [5]). Most of these procedures have emphasized automated reasoning over goal models, placing more value in the results of the analysis than in the interactive process of analyzing and exploring the model. However, our experience has shown that the informal and incomplete nature of goal models used for both Enterprise Modeling and Early Requirement Analysis is better served by interactive, qualitative analysis, allowing for the use of domain-specific knowledge to compensate for model incompleteness, and allowing for an interactive process of inquiry and questioning concerning the model domain.

We introduce a qualitative, interactive evaluation procedure for goal- and agent-oriented models, allowing the user to compare alternatives in the domain, asking "what if?"-type questions. Alternatives can include alternative system or process design choices, or alternative courses of actions, capabilities, and commitments. We also introduce a sample methodology using this procedure to guide users through the process of modeling and evaluation. As goal models are often created manually as informal sketches, it is important for analysis procedures and methodologies to be easy to apply. We present the procedure informally, using prose, to facilitate easy understanding and manual application. Although the procedure has now been implemented in the open-source, Eclipse-based OpenOME tool [6], past case studies involved manual application of the procedure. The procedure is presented in terms of the i* Framework; however, the procedure could be applied to other goal- and agent-oriented models, such as those created using the NFR Framework [7] or GRL [3].

The procedure and variations of the sample methodology have been tested via application to case studies, including a long-term project involving a large social service application as summarized in [8], [9], and [10] and an analysis of the intentions behind controversial new technology [11]. Our experience shows that in addition to helping compare alternatives, the analysis facilitates iteration in the modeling process, resulting in an overall improvement in understanding of the model and domain. We have developed and administered an exploratory experiment involving the evaluation procedure in an attempt to test the benefits discovered through case study application.

The procedure introduced in this work expands on a procedure introduced in the NFR Framework [7]. A short description of the procedure in this paper appears in [12]. This paper is organized as follows: Section 2 describes the sample methodology, including a short description of the i* Framework, Section 3 describes the evaluation procedure introduced in this work, Section 4 provides case study examples of the benefits of i* evaluation, Section 5 describes the experiment and its results, Section 6 describes related work, while Section 7 contains discussion, conclusions, and future work.

2 Modeling and Analysis with Goal- and Agent-Oriented Models: A Sample Methodology

Goal- and agent-oriented modeling frameworks, such as i*, are general enough to be used in several contexts, for modeling of the general enterprise (see for example [8]), modeling for early or later-stage system development ([1] or [4]), modeling for knowledge management [9], modeling for process redesign [13], and so on. However, a general methodology can be described, including model analysis, which can be applicable for modeling in multiple contexts. Because of the variety in the context and aims of such types of modeling activities, we advocate this methodology as only a general guide, or series of suggestions. Depending on the context, the role of stakeholders, and the specific required outcome of the modeling process, the methodology can be adapted as needed. The central themes of the methodology are incremental model development with analysis and iteration over models.

Our experience with creating models has indicated that the process of modeling and analysis is as important, perhaps even more important, for understanding and discovery as the resulting models. Ideally, this approach would be applied in cooperation with domain representatives. This allows representatives to have a sense of ownership over the model and the decisions made as a result of the modeling process, as described in [14]. However, it may be difficult to acquire stakeholder buy-in to the modeling process, and in these cases analysts can undertake the modeling process using other sources, including interviews, documents and observations.

As we use i* as an example goal- and agent-oriented framework, a basic knowledge of i* syntax is helpful in understanding the example methodology. The i* Framework facilitates exploration of an enterprise with an emphasis on social aspects by providing a graphical depiction of system actors, intentions, dependencies, responsibilities, and alternatives [1]. The social aspect of i* is represented by *actors*, including *agents* and *roles*, and the associations between them, (is-a, part-of, plays, covers, occupies, instantiates), which can be represented in an Actor Association (AA) model. Actors depend upon each other for the accomplishment of *tasks*, the provision of *resources*, the satisfaction of *goals* and *softgoals*. *Softgoals* are goals without clear-cut criteria for satisfaction. Dependencies between actors are represented in Strategic Dependency (SD) models. Actors can be "opened-up" in Strategic Rationale (SR) models using *actor boundaries* containing the *intentions* of an actor: desired goals and softgoals, tasks to be performed, and resources available. The interrelationships between intentions inside an actor are depicted with *Decomposition* links, showing the elements which are necessary in order to accomplish a task; *Means-Ends* links, showing the alternative tasks which can accomplish a goal; and *Contribution* links, showing the effects of softgoals, goals, and tasks on softgoals. Positive/negative contributions representing evidence which is sufficient enough to satisfy/deny a softgoal are represented by *Make*/*Break* links, respectively. Contributions with positive/negative evidence that is not sufficient to satisfy/deny a softgoal are represented by *Help*/*Hurt* links.

Although we present the six steps of our example methodology in a sequence, each step will often lead to changes in the results of previous steps. If the methodology is followed without the direct participation of stakeholders, each stage may result in questions which should be answered by domain experts. This knowledge should be incorporated back into the model at any stage. We will illustrate the method using a simplified example from the first phase of the youth counseling case study described in the introduction, selected results from this phase of the study are described in [8].

1. **Identify scope or purpose of the modeling process.** It is important to identify one or more issues of focus for the modeling process. This determines the scope of the analysis in each of the modeling steps, continually questioning the relevance of including certain actors, dependencies and intentions.
   Example: In the social service example, the purpose of the first phase of the study was to identify and evaluate the effectiveness of various technical alternatives for providing online youth counseling.
2. **Identify model sources.** As stated, ideally the models would be created along with selected domain stakeholders. Alternatively, interviews, enterprise documents, observations or other sources can be used.
   Example: In the example, stakeholders were generally unfamiliar with modeling as a tool for analysis and had difficulty committing significant amounts of time. As a result, models were developed by the analysts using stakeholder interviews and site visits.
3. **Identify relevant actors and associations.** With the model scope in mind, identify relevant enterprise actors and the relationships between them. This could include specific stakeholders or more abstract roles or organizations. Helpful analysis questions include: “Who is involved?” and “How are they related?”.
   Example: The actual case study identified 63 relevant actors. For our simplified example we will focus on youth, counselors and the counseling organization.
4. **Identify relevant dependencies.** In the same or a separate model, identify the dependencies between actors. Helpful analysis questions include: “Who needs what?” and “What do they provide in return?”.
   Example: The actual case study identified 405 potentially relevant dependencies, a subset of these dependencies are depicted in Fig. 1. To save space we have shown only the SR model, which includes the actors in the AA model and the dependencies in the SD model.
5. **Identify actor intentions.** This stage is divided into three iterative sub-steps:

   a. **Identify actor intentions:** Using the sources, identify what actors want, what tasks they perform, how they achieve things.

b. **Match dependencies to actor intentions:** Using the dependencies found in Stage 4, answer "why?" and "how?" questions for each dependency, linking all dependencies to existing or new intentions within an actor.

c. **Identify relationships between intentions:** Identify how the actor intentions relate to each other, whether it is through a functional AND/OR hierarchy or through positive or negative contributions. New intentions may be discovered. Ideally, no intentions should be isolated.

Example: A subset of the intentional elements identified in the case study is shown in Fig.1. This model captures two alternative ways to provide counseling services: text messaging through a cell phone and an online Cybercafe/ Portal/Chat Room. The effects of each option on the goals of each of example actor are captured via contribution links to softgoals. Even for this simplified example, a complex web of contributions and dependencies are formed.

6. **Evaluate alternatives within in model.** Apply the evaluation procedure introduced in this work, described in more detail in the next section.

The first application of the model typically involves evaluating the most obvious alternative, and often helps to test the "sanity" of the model. Isolated intentions which do not receive an evaluation value can be identified.

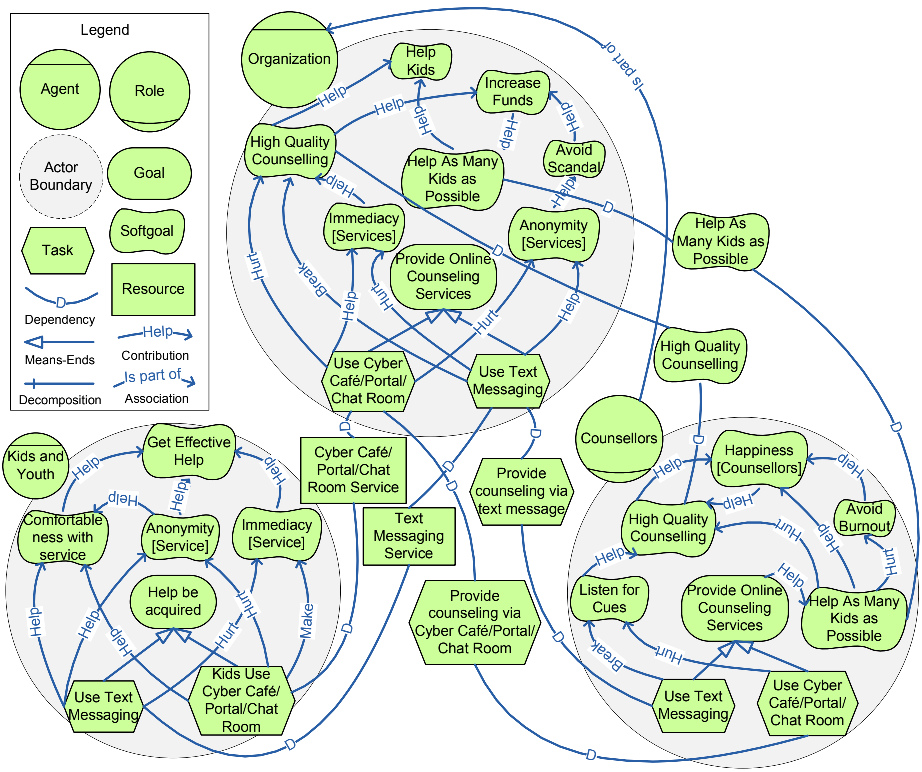

Fig. 1. Simplified Youth Counselling SR Example

Evaluation results which are not sensible can either reveal a problem in the model or an interesting discovery concerning the domain. Changes prompted by the evaluation results should be made in the model.

As the model evolves, more complicated or less obvious questions or alternatives can be analyzed. Further model changes can be made. The process continues until all viable alternatives are analyzed, an alternative has been selected, or a sufficient knowledge of the enterprise has been gained, depending on the initial purpose of the modeling process determined in Step 1.

Example: An example evaluation for the case study is presented in the next section as a means to illustrate the evaluation procedure. In the case study, several online counseling alternatives such as moderated forums, chats, email, and text messaging were analyzed and compared using the evaluation procedure.

3 A Qualitative, Interactive Evaluation Procedure for the i* Framework

Procedure Overview. The proposed procedure starts with an analysis question of the general form "How effective is an alternative with respect to model goals?". The procedure makes use of a set of qualitative evaluation labels, assigned to intentions to express their degree of satisfaction or denial. The process starts by assigning labels representing satisfaction and denial to intentions related to the analysis question. These values are propagated through the model links using defined rules. The interactive nature of the procedure comes when human judgment is used to combine multiple conflicting or partial values to determine the satisfaction or denial of a softgoal. The final satisfaction and denial values for the intentions of each actor are analyzed in light of the original question. An assessment is made as to whether the design choice is satisficed ("good enough"), stimulating further analysis and potential model refinement. More detail concerning the procedure can be found in [15].

Detailed Steps. We describe the steps of the evaluation procedure, followed by an explanation of the required concepts.

1. **Initiation:** The evaluator decides on an alternative and applies the initial evaluation labels to the model. The initial values are added to a label queue.

Iteratively, until the label queue is empty or a cycle is found:

2. **Propagation:** The evaluation labels in the label queue are propagated through all outgoing adjacent model links. Resulting labels propagated through non-contribution links are placed in the label queue. Results propagated through contribution links are placed into a "label bag" for that element.
3. **Softgoal Resolution:** Label bags are resolved by applying automatic cases or manual judgments, producing a result label which is added to the label queue.

4. **Analysis:** The final results are examined to find the impact of alternatives on stakeholder goals. Model issues can be discovered, further alternatives are evaluated.

Note that the procedure assumes that models are well-formed as per the syntax in [1]; however, as propagation is dependent on link type, most models can be evaluated.

Qualitative Evaluation Labels. We adopt the qualitative labels used in NFR evaluation, shown in Table 1. The *(Partially) Satisfied* label represents the presence of evidence which is *(insufficient)* sufficient to satisfy an intention. *Partially denied* and *denied* have the same definition with respect to negative evidence. *Conflict* indicates the presence of positive and negative evidence of roughly the same strength. *Unknown* represents the presence of evidence with an unknown effect. We use the "None" label to indicate a lack of any label. We use partial labels for tasks, resources, and goals, despite their clear-cut nature, to allow for greater expressiveness.

Initial Evaluation Values. In order to start an evaluation of a model, a set of initial values must be placed on the model, reflecting an analysis question and comprising Step 1 of the procedure. For example, in Fig. 1, if we wanted to ask "What is the effect of using a Cybercafe/Portal/Chat Room?", we would place initial values as shown in Fig 2 (circled labels).

Evaluation Propagation Rules. We define rules in order to facilitate a standard propagation of values given a link type and contributing label in Step 2 of the procedure. The nature of a *Dependency* indicates that if the element depended upon (*dependee*) is satisfied then the element depended for (*dependum*) and element depending on (*depender*) will be satisfied.

Decomposition links depict the elements necessary to accomplish a task, indicating the use of an AND relationship, selecting the "minimum" value amongst all of the values. Similarly, *Means-Ends* links depicts the alternative tasks which are able to satisfy a goal, indicating an OR relationship, taking the maximum values of intentions in the relation. To increase flexibility, the OR is interpreted to be inclusive. We expand the order of the values presented in the NFR Framework to allow for partial values, producing: None < ✗ < ✗• < ? < ⚡ < ✓• < ✓.

We adopt the *Contribution* link propagation rules from the NFR procedure, as shown in Table 1. These rules intuitively reflect the semantics of contribution links. Note that the "None" label is not propagated or placed in the label queue.

Resolving Multiple Contributions. Softgoals are often recipients of multiple contribution links. We adopt the notion of a "Label Bag" from [7], used to store all incoming labels for a softgoal. Labels in the label bag are resolved into a single label in Step 3, either by identifying cases where the label can be determined without judgment (Table 2), or by human judgment. For example, in Fig. 2, the Immediacy [Service]

Table 1. Propagation Rules Showing Resulting Labels for *Contribution* Links

Source Label		***Contribution* Link Type**						
	Name	**Make**	**Help**	**Some+**	**Break**	**Hurt**	**Some-**	**Unkn.**
✓	Satisfied	✓	✓•	✓•	✗	✗•	✗•	?
✓•	Partially Satisfied	✓•	✓•	✓•	✗•	✗•	✗•	?
⚡	Conflict	⚡	⚡	⚡	⚡	⚡	⚡	?
?	Unknown	?	?	?	?	?	?	?
✗•	Partially Denied	✗•	✗•	✗•	✓•	✓•	✓•	?
✗	Denied	✗	✗•	✗•	✓•	✓•	✓•	?

Table 2. Cases where Overall Softgoal Labels can be Automatically Determined

Label Bag Contents		Resulting Label
1.	The bag has only one label. Ex: {✗} or {✓•}	the label: ✗ or ✓•
2.	The bag has multiple full labels of the same polarity, and no other labels. Ex: {✓, ✓, ✓} or {✗, ✗}	the full label: ✓ or ✗
3.	All labels in the bag are of the same polarity, and a full label is present. Ex: {✓•, ✓, ✓•} or {✗, ✗•}	the full label: ✓ or ✗
4.	The human judgment situation has already occurred for this element and the answer is known	the known answer
5.	A previous human judgment situation for this element produced ✓ or ✗, and the new contribution is of the same polarity	the full label: ✓ or ✗

softgoal in Kids and Youth receives a satisfied and a partially satisfied label from incoming contributions links, resolved to a satisfied label using Case 3 in Table 2, reflecting the idea that evidence propagated to softgoals is roughly cumulative.

Human Judgment in Evaluation. Human judgment is used to decide on a label for softgoals in Step 3 for the cases not covered in Table 2. Human judgment may be as simple as promoting partial values to a full value, or may involve combining many sources of conflicting evidence. When making judgments, domain knowledge related to the destination and source intentions should be used. In this way, we compensate for the inherent incompleteness of social models. Areas where human judgment is needed can be considered for further model expansion; however, given the tradeoff between completeness and model complexity, it may not be feasible to altogether avoid human intervention for a particular model.

For example, the resulting label for Happiness [Counselors] in Fig. 2 is determined by human judgment. This softgoal receives partially denied labels from Avoid Burnout and High Quality Counseling, but receives a partially satisfied label from Help as many Kids as Possible, according to the propagation rules in Table 1. Here, using our knowledge of the domain, we decide that Counselors would be mostly unhappy, labeling the softgoal as partially denied. Situations such as this would be good areas for potential discussions with stakeholders involved in the modeling process.

Combinations of Links. Intentions in i* are often the destination of more than one type of link. Following strict i* syntax, this occurs when an element is the recipient of a dependency link and a means-ends/decomposition link or a contribution link. "Hard" links (Decomposition, Means-Ends and Dependency) are combined using an AND of the final results of each link type. If Contribution and Dependency links share the same destination, the result of the Dependency links are treated as a Make contribution, considered with the other contributions in the label bag. An example of this type can be seen in High Quality Counseling in the Organization.

Incomplete Labels. In the procedure, information present in each step is propagated, even if this information in incomplete, i.e., other incoming contributions are missing. As a result, the evaluation labels for an element may change throughout the procedure and the same softgoal may require human judgment multiple times.

Detecting Cycles. Goal models often contain cycles, values which indirectly contribute to themselves and may cause fluctuating values. Our implementation of the procedure places a cap on the number of value fluctuations possible for an intention. Experience has shown that during manual application of the procedure the presence of cycles becomes apparent to the evaluator after a few iterations. We recommend that the evaluator manually selects appropriate converging values, marking the cycle as an area which may need further model refinement.

Example Evaluation. We return to the question posed in Section 2 concerning Fig. 2, "What is the effect of using a Cybercafe/Portal/Chat Room?". Results can be analyzed from the point of view of each actor. For Kids and Youth, the Cybercafe/Portal/Chat Room provides Immediacy as well as a Comfortable Service, but jeopardizes Anonymity, making the overall assessment weakly satisfied for Get Effective Help. From the point of view of Counsellors, the alternative has a positive effect on Help as Many Kids as Possible, but has a negative effect on Burnout and the Quality of Counselling, making the overall assessment to Counsellor Happiness weakly negative. From the point of view of the organization, the service also has a positive effect on Helping as Many Kids and Possible and Immediacy, but has a negative effect on Anonymity, Avoiding Scandal, Increasing Funds, and the Quality of Counselling. There is conflicting evidence for the

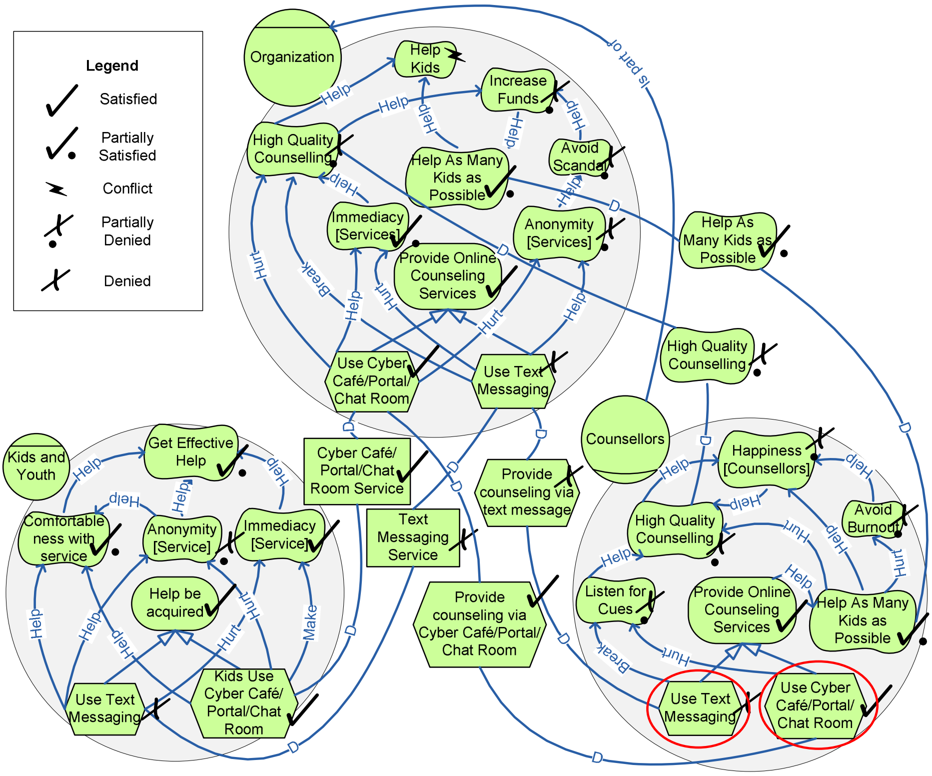

Fig. 2. Simplified Youth Counselling Example showing Final Evaluation Results

ability to Help Kids. Overall, this alternative is judged to be not viable. A further round of evaluation is needed to assess the other alternative in the model, text messaging, and to use the goals in the model to brainstorm further online counselling services which balance concerns more effectively.

4 Experience from Case Studies

We have applied our procedure and methodology to several case studies involving analysis of socio-technical settings, two of which are described here.

In the case study involving a large social service organization, [8], the evaluation procedure was applied in several stages. The first stage of the project is described in Section 2 as an illustrative application of our example methodology. Here the procedure was applied manually to large models (the largest had 353 intentions) in order to analyze and compare the effectiveness of technology options for providing counseling over the internet. The results were presented to the organization using reports and presentation slides containing small excerpts of the model. The analysis was well-received by the organization, bringing to light several issues and provoking interesting discussion. However, due partially to a lack of resources available to handle online counseling traffic, the organization opted to continue to use a modified version of the moderated bulletin board option already in place.

The next stage of the project focused on increasing the efficiency of the existing system. The evaluation procedure was used to analyze various configurations of a moderated bulletin board system, with feedback from the stakeholders used to validate the findings. The final outcome was a requirements specification document provided to the organization. Due to resource limitations and the risks involved in deploying a new system, the organization opted to modify their existing system instead of implementing a new system based on the specification.

A later phase of the project with the same organization focused on applying enterprise modeling to analyze the knowledge management needs of the organization [9]. The evaluation procedure was applied manually to large models in order to evaluate the situational effectiveness of a variety of technologies for storing and distributing knowledge, including wikis and discussions forums. It was discovered, for example, that the features of a wiki were not effective in satisfying the goals of the organization, while a discussion forum, with a set of specific features, showed more promise. We found the procedure to be effective in facilitating a comparison between technologies, with the results reported back to the organization in a series of reports, receiving positive stakeholder feedback. The largest model evaluated in the study contained 544 elements, helping to demonstrate the scalability of the procedure.

We used the opportunity presented by the case study to test the application of model patterns to i* modeling [10]. Here, general models representing technologies were integrated into context-specific models describing the organization. In this case, the patterns and the situational models involved underwent evaluation, using the models to answer various interesting questions, before integration. Our observations in this and other application of the procedure attest to the model iteration provoked by evaluation. For example, before evaluation in the pattern study a context-specific model focusing on communication contained 181 links and 166 elements, while after evaluation the same model had 222 links and 178 elements, a difference of 41 and 12 respectively. In another example, the link count rose from 59 to 96 and the element count rose from 59 to 76. These numbers do not take into account changes such as moving links or changing element names. Models in this study were created by three individuals with evaluation performed by two individuals, helping to demonstrate that this effect is not specific to a particular modeler or evaluator.

Our experience has shown that analysis can also be used as a means of understanding, justifying and explaining complex situations. Examples of this type can be found in a further case study, describe in [11], where evaluation is used to describe the motivations behind stakeholders involved in Trusted Computing (TC). Here, evaluation was used to help demonstrate the differences between proponents and opponents of Trusted Computing Technology, with proponents claiming it help to ensure security for the user, and opponents claiming the technology provided less security and more restrictions by enforcing Digital Right Management. The evaluation procedure helped to show the effects of these different perceptions on the goals of participating actors such as Technology Producers, License/Copyright Owners, Technology Consumers, and Malicious Parties, even when these actors and their goals were not directly connected to the differing effects of TC technology.

Our case study experience demonstrated the ability of the procedure to provoke further elicitation and subsequent model iteration. For example, in the TC case study, although the model appeared to be sufficiently complete, one of the first rounds of analysis of the TC Opponent point of view revealed that Technology Users would not buy TC Technology. Although this may be the case for some Users, obviously the makers of TC Technology envisioned some way in which users would accept their product. These results led the modeler to further investigate the sources, including factors such as product lock-in, more accurately reflecting the domain.

Prompted by our case study experience, we developed and carried out an exploratory experiment designed to test some of the perceived benefits of the procedure, described in the next section.

5 Experimental Results

Observations in case studies have shown that the evaluation procedure described in this work aids in finding non-obvious answers to analysis questions, prompts improvements in the model, leads to further elicitation, and leads to a better understanding of the domain. Our experiment begins to test whether these effects are specific to our procedure or are a product of any detailed examination of a model.

The experimental models were taken from a study applying goal-oriented analysis to the sustainability issues for the ICSE conference [16]. The study produced a series of models focusing on actors in the domain of conference planning. For the experimental investigation, the five participants of that study, including one of the authors, were asked to evaluate two different questions over three models, once without using the procedure and, after training, once using the procedure. The results were compared in terms of analysis findings, questions discovered, model changes, and time taken. The three models contained between 36 and 79 intentions, 50 and 130 links, and 5 and 15 actors. Participants were given two non-trivial analysis questions related to goal satisfaction specific to each of the three models. Participants were asked to answer follow-up questions: Did model changes improve the model quality? Do you have a better understanding of the model and domain? Did this increase more or less, with or without using the procedure? Would you use the procedure again?

We examine several aspects of the results. First the differences in analysis results not using, and then using the procedure, helping to show that the procedure finds non-obvious analysis answers. We observe that the participants made a total of 40 changes to their analysis results after applying the procedure and that changes were made for each question over each model. All participants made changes to their analysis results, with each participant making between 7 and 11 total changes in all question-model combinations. A breakdown of the types of changes is omitted due to space restrictions.

Next, we count the changes made to the models not using and using the procedure. Overall, in evaluating two questions over three models, the 5 participants made a total of 71 changes without using the procedure and then 40 changes using the procedure. Changes were made for each model, and all participants made changes. These results may indicate that the iteration provoked by the procedure may have more to do with forcing the user to carefully manually examine the model than with the procedure itself. However, we note that the participants found 40 additional changes using the procedure to answer the questions for the second time.

In examining the model quality improvement, three out of five participants said that changes made to the models improved the quality of the model. These participants indicated the quality was improved through changes made both with and without the procedure. The other two participants did not feel they had made significant changes to the models in either stage, with one stating that "additional knowledge information would be needed to really improve the quality of the models", and the other echoing the sentiment. These results help to emphasize the incomplete and iterative nature of such models, and their ability to prompt further elicitation. Along this line, we observe that participants came up with between 5 and 16 questions each, at total of 26 questions were derived without using the procedure, while an additional 19 were derived while using the procedure, for a total of 45. Although many questions were derived without using the procedure, we observe that application of the procedure provoked a number of further questions, even though the same analysis questions were being evaluated.

All five participants reported a better understanding of the domain after this exercise, with all participants claiming that they gained a better understanding using the evaluation procedure than using no procedure. The average time to answer a question without the procedure was 9.5 minutes (standard deviation of 4.6) compared to 11.1 minutes (standard deviation of 6.0) using the procedure. Although the variance is high, we see that working with procedure takes only slightly more time than without. Finally, all five participants said they would use the procedure again if they had to evaluate another i* model.

6 Related Work

Goal concepts are prominent in a number of modeling frameworks, notably in "goal-oriented" requirements engineering (e.g. [17] and [18]) as well as in enterprise modeling e.g. [14], [19]. While all of these frameworks provide for the representation of goals and relationships among goals, only some of the frameworks have associated procedures for determining whether goals are met, for example [2], [3], [4], and [5]. Most of these procedures have taken a more formal, automated, or quantitative approach to goal model analysis. We argue that such procedures are more suitable later in the analysis, when more complete and detailed system information is available, and where models are more stable and appropriate for automated reasoning. The interactive, qualitative approach, such as the one introduced in this work, is more appropriate for early analysis, to gain a high-level understanding of the domain, and to discover and evaluate alternatives with stakeholders. Once the number of alternatives has been narrowed using interactive, qualitative evaluation, more detailed information can be added to the model and various forms of quantitative or automated analysis can be applied in order to further test the feasibility of a particular alternative.

An interactive qualitative evaluation procedure based on the notion of goal "satisficing" was first introduced to evaluate Softgoal Interdependency Graphs as part of the NFR Framework [7]. Previous work has used this procedure evaluate i* models, (see [14] for example), assuming that the NFR procedure could be easily extended for use with i*, without describing the necessary extensions, modifications, or additional benefits. Application of the NFR procedure to i* models in case studies such as [11] has shown that the level of interactivity is too restrictive, assigning a conflict label to all goals with conflicting evidence. We build upon this earlier procedure by introducing aspects which cover agent-oriented concepts, providing steps for application, adjusting the use of human intervention and more thoroughly exploring issues such as initial values and convergence.

Alternative methodologies to direct the creation of i* models have been introduced. The RESCUE method, aimed for system design or redesign, directs the development of several streams of models in parallel including i*, activity, use case, and requirements models [21]. The Process Reengineering i* Method (PRiM) builds on this approach, constructing i* models to understand and redesign business processes and associated information systems [13]. The methodology introduced in Section 2 is more general, applicable to modeling aspects of an enterprise which may or may not be specific to an information system or to actors involved in a particular process.

7 Discussion, Conclusions, and Future Work

In this work, we have identified the need for systematic evaluation of alternatives within models capturing the goals of an enterprise. We have introduced a simple procedure which builds on the NFR procedure, expanding the procedure to deal with agent-specific constructs, and more thoroughly exploring issues such as initial values, propagation rules, and human judgment. A sample methodology describing how to use this procedure in the process of enterprise modeling has been presented. We have explored the benefits the methodology and evaluation procedure, including analysis, model iteration, and elicitation by describing application to two case studies.

Experience has shown that it is difficult to acquire stakeholder buy-in to the modeling process, often due to the considerable time taken by the process or unfamiliarity with modeling as an analysis tool. Existing case studies have involved modeling and evaluation by analysts using interviews, documents, or site observations. Although this process is very useful to help the analysts understand and explore the domain, it is difficult to fully present or validate the resulting models and the results of evaluation. While the analysts who have constructed the model and performed the evaluations are able to understand the model and evaluation results, the models are too large and the evaluation results are too complicated to be easily understood by stakeholders. Thus far, we have only investigated model evaluation in the context of a single modeler. Future work should investigate its role in collaborative or group settings. Although experimental results provide some confidence in the ability of users to learn and apply evaluation, participatory studies would help to confirm the ability of domain users to apply the procedure on their own. Such studies would also help to further assess the mechanism of evaluation, including the appropriateness of propagation rules.

Results of our exploratory experiment indicate that the evaluation procedure prompts changes to evaluation results and may prompt model iteration and elicitation beyond analysis without a systematic procedure. The participants have reported that the procedure provides a better understanding of the model and domain. However, the experiment suffers from several threats to validity, including the small number of participants. Using the lessons learned from this experiment, we hope to conduct further experiments with more participants. Future experiments should try to push the limits of evaluation without a systematic procedure by asking participants to examine the model multiple times. Further studies can explore the perceived benefits of the applying the procedure, including studies to determine whether these benefits are specific to the qualitative, interactive procedure introduced in this work, or apply more generally to other Goal- and Agent-Oriented evaluation procedures.

In order to make our description of the evaluation procedure more concrete, we have applied it to the i* Framework, potentially limiting applicability. However, as most other Goal- and/or Agent-Oriented Frameworks, such as the NFR Framework [7] or GRL [3], are syntactic subsets of i*, our procedure can be easily extended to other, similar frameworks. Future work could include adapting and applying the evaluation procedure described in this work for use with other goal modeling frameworks, such as the goal component in EKD models.

The procedure introduced in this work can be expanded in several ways, for example: capturing the rationale and assumptions behind human judgments, evaluating the satisfaction of actors as in [3], expanding analysis in a top-down direction as explored in [22], allowing for constraints as in [2], facilitating the traceability of evidence, and giving users selection over different qualitative scales.

Acknowledgements. Financial support has been provided by the Natural Sciences and Engineering Research Council of Canada, Bell University Laboratories, and the Ontario Graduate Scholarship Program.

References

1. Yu, E.: Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering. In: 3rd IEEE International Symposium on Requirements Engineering (RE 1997), pp. 226–235. IEEE Press, New York (1997)
2. Giorgini, P., Mylopoulos, J., Sebastiani, R.: Simple and Minimum-Cost Satisfiability for Goal Models. In: Persson, A., Stirna, J. (eds.) CAiSE 2004. LNCS, vol. 3084, pp. 20–35. Springer, Heidelberg (2004)
3. Amyot, D., Ghanavati, S., Horkoff, J., Mussbacher, G., Peyton, L., Yu, E.: Evaluating Goal Models within the Goal-oriented Requirement Language. Int. Journal of Intelligent Systems (IJIS) (to appear)
4. Letier, E., van Lamsweerde, A.: Reasoning about Partial Goal Satisfaction for Requirements and Design Engineering. In: 12th ACM International Symposium on the Foundations of Software Engineering (FSE 2004), pp. 53–62. ACM, New York (2004)
5. Franch, X.: On the Quantitative Analysis of Agent-Oriented Models. In: Dubois, E., Pohl, K. (eds.) CAiSE 2006. LNCS, vol. 4001, pp. 495–509. Springer, Heidelberg (2006)
6. OpenOME, `https://se.cs.toronto.edu/trac/ome/wiki`
7. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: Non-Functional Requirements in Software Engineering. Kluwer Academic Publishers, Norwell (2000)
8. Easterbrook, S.M., Yu, E., Aranda, J., Fan, Y., Horkoff, J., Leica, M., Qadir, R.A.: Do Viewpoints Lead to Better Conceptual Models? An Exploratory Case Study. In: 13th IEEE International Requirements Engineering Conference (RE 2005), pp. 199–208. IEEE Press, New York (2005)
9. Strohmaier, M., Yu, E., Horkoff, J., Aranda, J., Easterbrook, S.: Analyzing Knowledge Transfer Effectiveness - An Agent-Oriented Approach. In: 40th Hawaii International Conference on Systems Science (HICSS-40), p. 188b. IEEE Press, New York (2007)
10. Strohmaier, M., Horkoff, J., Yu, E., Aranda, J., Easterbrook, S.: Can Patterns improve i* Modeling? Two Exploratory Studies. In: Paech, B., Rolland, C. (eds.) REFSQ 2008. LNCS, vol. 5025, pp. 153–167. Springer, Heidelberg (2008)
11. Horkoff, J., Yu, E., Liu, L.: Analyzing Trust in Technology Strategies. In: International Conference on Privacy, Security and Trust (PST 2006), pp. 21–32 (2006)
12. Horkoff, J., Yu, E.: A Qualitative, Interactive Evaluation Procedure for Goal- and Agent-Oriented Models. In: CAiSE Forum. CEUR Workshop Proceedings (2009)
13. Grau, G., Franch, X., Maiden, N.A.M.: PRiM: an i*-based process reengineering method for information systems specification. Information and Soft. Tech. 50(1-2), 76–100 (2008)
14. Stirna, J., Persson, A.: Ten Years Plus with EKD: Reflections from Using an Enterprise Modeling Method in Practice. In: Pernici, B., Gulla, J.A. (eds.) Proceedings of the 11th International Workshop on Exploring Modeling Methods in Systems Analysis and Design (EMMSAD 2007), pp. 99–108. CEUR-WS.org (2007)
15. Horkoff, J.: An Evaluation Algorithm for the i* Framework. Master's Thesis, Department of Computer Science, University of Toronto (2006)
16. Cabot, J., Easterbrook, S., Horkoff, M.J., Lessard, L., Liaskos, S.: Integrating Sustainability in Decision-Making Processes: A Modelling Strategy. In: ICSE 2009 New Ideas and Emerging Results, NIER 2009 (2009)
17. van Lamsweerde, A.: Goal-Oriented Requirements Engineering: A Guided Tour. In: 5th IEEE International Symposium on Requirements Engineering (RE 2001) (invited paper), pp. 249–263. IEEE Press, New York (2001)

18. Kavakli, E., Loucopoulos, P.: Goal Driven Requirements Engineering: Analysis and Critique of Current Methods. In: Krogstie, J., Halpin, T., Siau, K. (eds.) Information Modeling Methods and Methodologies, pp. 102–124. Idea Group (2004)
19. Rolland, C., Prakash, N.: Bridging the Gap Between Organisational Needs and ERP Functionality. Requirements Engineering 5(3), 180–193 (2000)
20. Liu, L., Yu, E.: Designing Information Systems in Social Context: A Goal and Scenario Modelling Approach. Information Systems 29(2), 187–203 (2004)
21. Maiden, N.A.M., Jones, S.V., Manning, S., Greenwood, J., Renou, L.: Model-Driven Requirements Engineering: Synchronising Models in an Air Traffic Management Case Study. In: Persson, A., Stirna, J. (eds.) CAiSE 2004. LNCS, vol. 3084, pp. 368–383. Springer, Heidelberg (2004)
22. Horkoff, J., Yu, E.: Qualitative, Interactive, Backward Analysis of i* Models. In: 3rd International i* Workshop, pp. 43–46. CEUR-WS.org (2008)