Interactive Analysis of Agent-Goal Models in Enterprise Modeling

Jennifer Horkoff*[1], Eric Yu[2]
[1]Department of Computer Science, [2]Faculty of Information,
University of Toronto, Canada

ABSTRACT
Understanding and analyzing the needs of an enterprise in the early stages of a project requires knowledge about stakeholders, their goals, interactions, and alternative actions. Agent-goal models offer a way to systematically and graphically capture this information, even as it evolves through continued elicitation. However, the complexity of resulting models makes it difficult to evaluate the achievement of key stakeholder goals within a model without applying systematic analysis procedures. Existing approaches to agent-goal model evaluation focus on automated procedures, without explicitly promoting model iteration and domain elicitation. We argue that "Early" Enterprise modeling requires analysis procedures which account for the incompleteness and informality of early agent-goal models, facilitating iteration, elicitation, and user participation. We introduce a qualitative, interactive evaluation procedure for agent-goal models, using the i* Framework in our illustrations. Case study experience and the results of an exploratory experiment show the applicability of the procedure to early enterprise analysis.

Keywords: Goal- and Agent-Oriented Models, Goal Modeling, Model Analysis, Enterprise Modeling, Information Systems Modeling, Requirements Modeling.

INTRODUCTION
The success of an enterprise is often affected by intangible, abstract goals which play a role in achieving customer, colleague, and employee satisfaction, and which typically depend on the actions of other parties. Goal- and Agent-Oriented Modeling frameworks, such as i* (Yu, 1997), have been introduced in order to model and explore socio-technical domains including actors or stakeholders, their goals and responsibilities, dependencies and alternatives. This approach was initially aimed as a first step in a system development process, as part of "Early" Requirements Engineering (RE), but it is also applicable to the early stages of modeling and understanding an enterprise, including its internal operations and relationships to the external environment. Such "Early" Enterprise Modeling (EM) can be used to understand the objectives of stakeholders, explore alternative courses of action, analyze their impacts, assess whether stakeholder objectives are met, make tradeoffs among competing goals, and facilitate organizational decision making.

Consider an example enterprise: a not-for-profit organization that provides counseling for youth over the phone, but must now also offer counseling via 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, balancing the needs of multiple parties?

Goal- and Agent-Oriented Models (agent-goal 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, when a model contains multiple, multi-step relationship links, it can be difficult to trace the effect of a particular alternative on the satisfaction of one or more goals. 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 customer satisfaction, 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.

A number of analysis procedures for analyzing goal models have been introduced (for example: Giorgini et al., 2004; Amyot et al., in press; Letier & van Lamsweerde, 2004; Franch, 2006). However, these procedures have focused on automated reasoning, placing more value in the results of the analysis than in the interactive process of analyzing and exploring the model. An ideal analysis procedure for Early RE or EM agent-goal models would consider the informal and incomplete nature of such models, would facilitate iteration and domain exploration, and would be simple to apply. Our experience has shown that interactive, qualitative evaluation allows for the use of domain-specific knowledge to compensate for model incompleteness, and encourages an interactive process of inquiry over the domain.

In this work 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 present the procedure informally, using prose, to facilitate easy understanding and optional manual application. We introduce a sample methodology using this procedure to guide users through the process of modeling and evaluation. Although the procedure has now been implemented in the open-source, Eclipse-based OpenOME tool (OpenOME, 2009), 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 agent-goal models, such as those created using the NFR Framework (Chung et al., 2000) or GRL (Amyot et al., in press).

The procedure and variations of the sample methodology have been tested in case studies, including an analysis of the intentions behind controversial new technology (Horkoff et al., 2006) and a long-term project involving a large social service application (summarized in Easterbrook et al., 2005; Strohmaier et al., 2007; Strohmaier et al., 2008). Case study experience shows that the procedure addresses many desired characteristics for Early EM, including the ability to provoke model iteration and domain exploration. We have administered an exploratory experiment applying the evaluation procedure to further test key benefits discovered in case studies.

The procedure introduced in this work expands on a procedure introduced in the NFR Framework (Chung et al., 2000). A short description of the procedure in this paper appears in Horkoff & Yu (2009a). This work is an expansion of Horkoff & Yu (2009b). The paper is organized as follows: Section 2 introduces agent-goal modeling, Section 3 motivates qualitative, interactive evaluation for agent-goal models, Section 4 describes a sample agent-goal model methodology, Section 5 outlines the evaluation procedure introduced in this work, Section 6 describes case study applications of the procedure, Section 7 outlines the experiment and its results, Section 8 contains a discussion, Section 9 describes related work, while Section 10 provides conclusions and outlines future work.

AGENT-GOAL MODELING

Agent-goal models are intended to facilitate exploration of the system domain with an emphasis on social aspects by providing a graphical depiction of system actors including their intentions, dependencies, responsibilities, alternatives and vulnerabilities (Yu, 1997). As we use i* as an example agent-goal framework, a basic knowledge of i* syntax is necessary.

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 intentional elements (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.

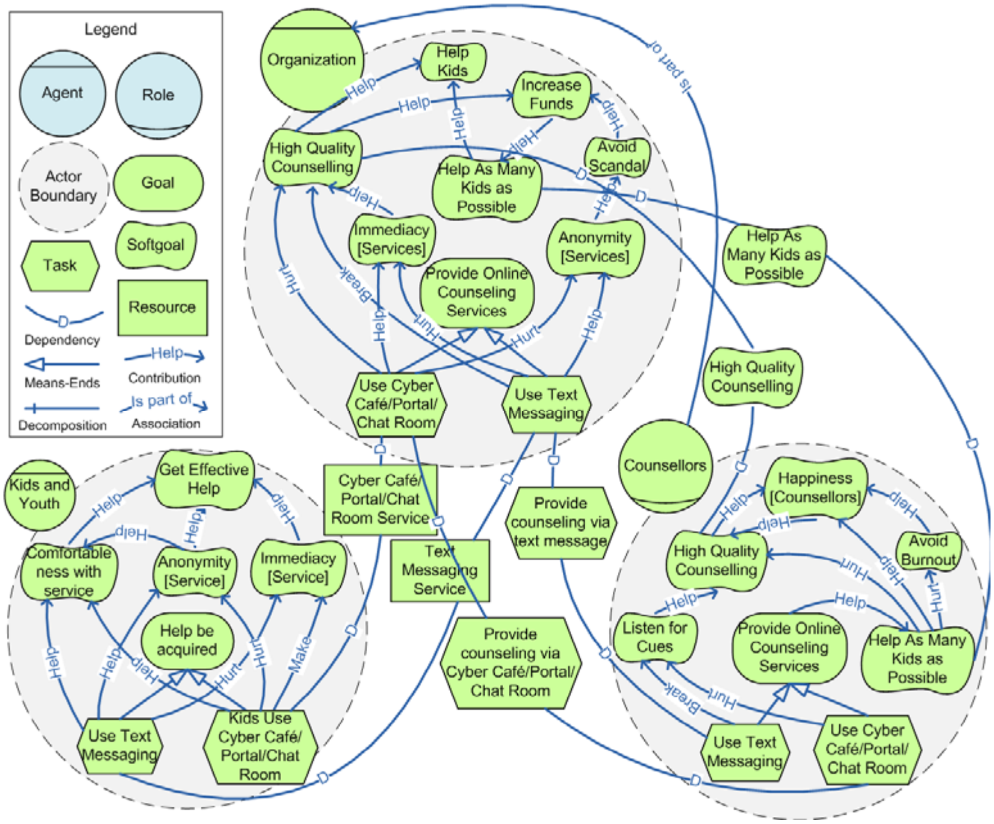

Figure 1. SR Model for Youth Counseling (Simplified)

Figure 1 contains an example i* model showing a simplified view of the first phase of the youth counseling case study described in the introduction. Although this model may seem complex at first, a reader can understand the model by examining it actor by actor and element by element. This model contains three actors: the Organization (top), Kids and Youth (bottom left), and Counselors (bottom right). The Organization, an agent, wants to achieve several softgoals, including Helping Kids, Increasing Funds, and providing High Quality Counseling. These goals are difficult to precisely define, yet are critical to the organization. The Organization as the "hard" goal of Providing Online Counseling Services and explores two alternative tasks for this goal: Use Text Messaging and Use Cyber Café/Portal/Chat Room. These alternatives contribute positively or negatively to various degrees to the Organization's goals, which in turn contribute to each other. For example, Use Text Message hurts Immediacy which helps High Quality Counseling.
The Organization depends on the Counselors to provide the alternative counseling services and for many of its softgoals, for example, High Quality Counseling. Kids and Youth depend on the Organization to provide various counseling services, such as Cyber Café/Portal/Chat Room. Both the Counselors and Kids have their own goals to achieve, also receiving contributions from the counseling alternatives. Although the internal goals of each actor may be similar, each actor is autonomous: High Quality Counseling may mean something different for the Counselor than for the Organization.

Examining the example model raises several questions: Which counseling alternative is the most effective? Effective for whom? Can each actor's goals be achieved? What information is missing? Although some questions may be answered by studying the model, tracing effects consistently without guidance quickly becomes complex. A systematic analysis procedure which accounts for the information and incomplete nature of the model is needed.

THE NEED FOR INTERACTIVE, QUALITATIVE ANALYSIS FOR EARLY AGENT-GOAL MODELS

Although a number of evaluation analysis procedures for agent-goal models have been introduced, when goal-oriented techniques are used, explicit analysis procedures are often not applied. For example, work by Maiden et al. (2004) uses synchronization with i* and other models to find omissions in the domain, but does not apply analysis to i* models individually. Work by Gordijn et al. (2006) uses comparisons of i*-like models to value models for a requirements analysis of enterprise networks without applying explicit analysis to goal models. We can consider several reasons for the frequent omission of explicit goal model analysis. Earlier presentations of agent-goal models have not emphasized model analysis (Yu, 1997). Most procedures introduced for agent-goal models provide automated reasoning, but when models are incomplete and imprecise, it is hard for users – analysts and stakeholders – to trust the results of fully automated analysis. Depending on the complexity and the transparency of the analysis procedure, it may be difficult to understand how results are achieved, making validation of results challenging. Because Early EM models are used for exploration and early decision making, the models often change and expand along with the analyst's perception of the enterprise. Many analysis procedures for agent-goal models are presented as a means of finding answers over static models, but analysis procedures for dynamic models can also be used as an exploratory tool to test the sanity of the model in its current stage.

An ideal analysis procedure for Early EM agent-goal models would consider the informal and incomplete nature of such models, would facilitate model iteration and domain exploration, and would offer transparency and simplicity. Such a procedure should encourage stakeholder validation, domain learning, and trust, both in the modeling process and the analysis results. Table 1 lists desired characteristics for agent-goal model analysis, including both general characteristics and those of greater concern to early, exploratory models. Although previously introduced analysis procedures have had success in addressing general desirable characteristics to agent-goal model analysis, we believe they fall short on characteristics more important in early analysis.

Table 1. Desired Characteristics for Agent-Goal Model Analysis

	General Characteristics		Early Analysis Characteristics
G1	Assess achievement of objectives	**E1**	Account for incompleteness
G2	Explore alternatives	**E2**	Account for informality
G3	Assist in making tradeoffs	**E3**	Facilitate iteration
G4	Evaluate over complex models	**E4**	Facilitate elicitation and exploration
G5	Consistent analysis	**E5**	Transparent and simple
		E6	Amenable to stakeholder validation
		E7	Stakeholders trust results

GUIDELINES FOR MODELING AND ANALYSIS WITH AGENT-GOAL MODELS

In order to facilitate the use of agent-goal models for Early EM, we provide a set of guidelines for model creation, iteration, and analysis. Case study experience has led to the belief that a highly specific methodology for creating and analyzing agent-goal models may be too restrictive, due to varying characteristics of the enterprise and available modelers. As a result, we advocate this methodology as only a general guide, or a 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.

Our experience with modeling has shown 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 by Stirna & Persson (2007). 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.

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.

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 modeling participants and/or model sources.** As stated, ideally the models would be created along with selected domain stakeholders who would act as a source for the information captured in the models. Alternatively, if stakeholders are not directly available, interviews, documents, observations or other sources can be used. These sources could also be used to supplement the knowledge of any participating stakeholders.
   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 information gained through 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. In our simplified example we 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 Figure 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 Figure 1. Even for this simplified example, a complex web of contributions and dependencies are formed.

6. **Evaluate alternatives within the 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. 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.

A QUALITATIVE, INTERACTIVE EVALUATION PROCEDURE FOR THE I* FRAMEWORK

Procedure Overview

The proposed procedure starts with an analysis question of the 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 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 Horkoff (2006).

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

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 described by Yu (1997); 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 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, a set of initial values reflecting an analysis question is placed on the model. For example, in Figure 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 Figure 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 the following ordering:

✗ < ✗• < ? < ⤥ < ✓• < ✓.

We adopt the *Contribution* link propagation rules from the NFR procedure, as shown in Table 2. These rules intuitively reflect the semantics of contribution links.

Resolving Multiple Contributions: Softgoals are often the recipient of multiple contribution links. We adopt the notion of a “Label Bag” from Chung et al. (2000), 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 Figure 2, the Immediacy [Service] 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 3, reflecting the idea that evidence propagated to softgoals is roughly cumulative.

Table 2. 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 3. 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 ✗

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 3. 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.

For example, the resulting label for Happiness [Counselors] in Figure 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 2. 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. 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, allowing the evaluator to select an appropriate converging value.

Example Evaluation: We evaluate one of the counseling options, asking "What is the effect of using a Cybercafe/Portal/Chat Room?" Results shown in Figure 2 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. From the point of view of the Organization, the service also has a positive effect on Helping as Many Kids as 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 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.

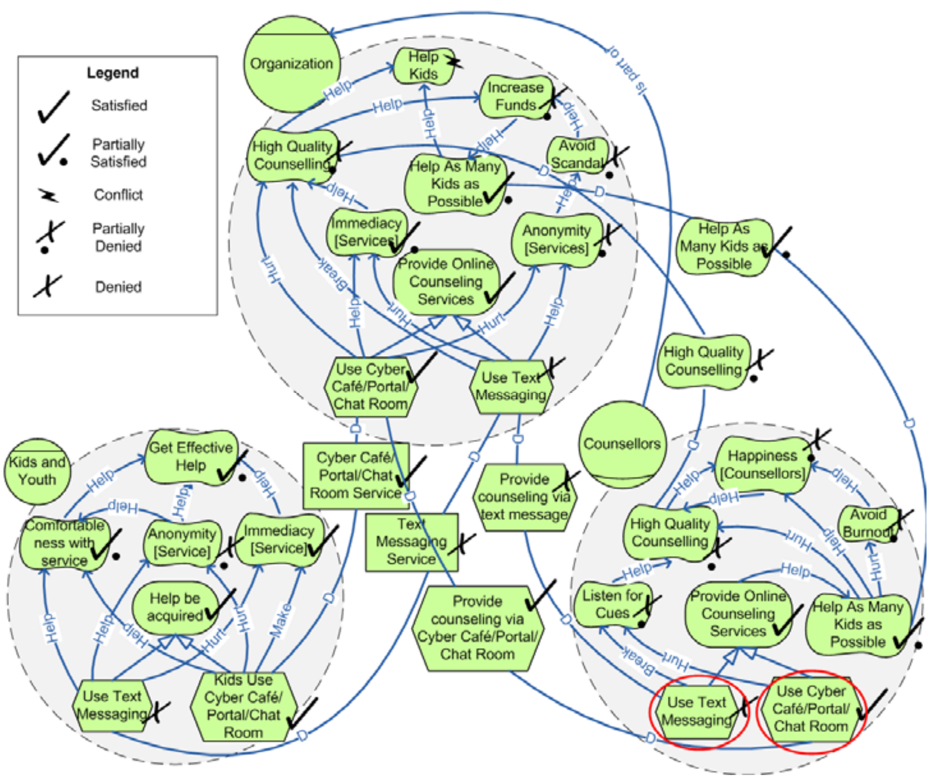

Figure 2. SR Model for Youth Counselling (Simplified) showing Final Evaluation Results

EXPERIENCE FROM CASE STUDIES

We have applied our procedure and methodology to several case studies involving analysis of socio-technical settings. These experiences provided useful lessons, both specific to agent-goal analysis and to modeling socio-technical domains in general.

Initial Exploratory Studies

Initially i* modeling was used to model and analyze domains with textual documents as sources. Models were created to analyze the Montreux Jazz Festival, based on business models created by Osterwalder (2004). Models were created to analyze E-Commerce data exchanges, as described by Spiekermann (2003), and Economic Information Security as described by Anderson (2001). Initially, these studies applied the qualitative, interactive evaluation procedure described as part of the NFR procedure, with adjustments and additions made for the agent-oriented concepts in i*. Experiences showed that this procedure did not provide enough flexibility due to too much automation. For example, all softgoals receiving conflicting evidence were automatically assigned a conflict label. The high-level and informal nature of the analysis domains called for a procedure with more user control. Adjustments to the role of human judgment were made, and a more precise description of the procedure was produced, as described by Horkoff (2006), producing an algorithm similar to the description in this work. The next steps involved testing the procedure on larger, more complex, case studies.

Lessons Learned: Qualitative, interactive evaluation was helpful in getting more value out of agent-goal models, especially concerning domain understanding and model improvement, but flexibility in combining conflicting evidence is needed.

Trusted Computing Case Study

In the next study, i* and qualitative interactive analysis were applied to describe the motivations behind stakeholders involved in Trusted Computing (TC), described in (Horkoff et al., 2006). Models were created based on publically available textual sources. Here, evaluation was used to help demonstrate how proponents and opponents of Trusted Computing Technology differed in their viewpoints. Proponents claimed it helps to ensure security for the user, while opponents claimed the technology provided less security and more restrictions by enforcing digital rights management. The evaluation procedure helped to show the effects of the different perceptions of TC 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. Figure 3 shows a high-level overview of the TC Opponent model, including evaluation.

The TC study demonstrated the ability of the procedure to provoke further elicitation and subsequent model iteration. For example, 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.

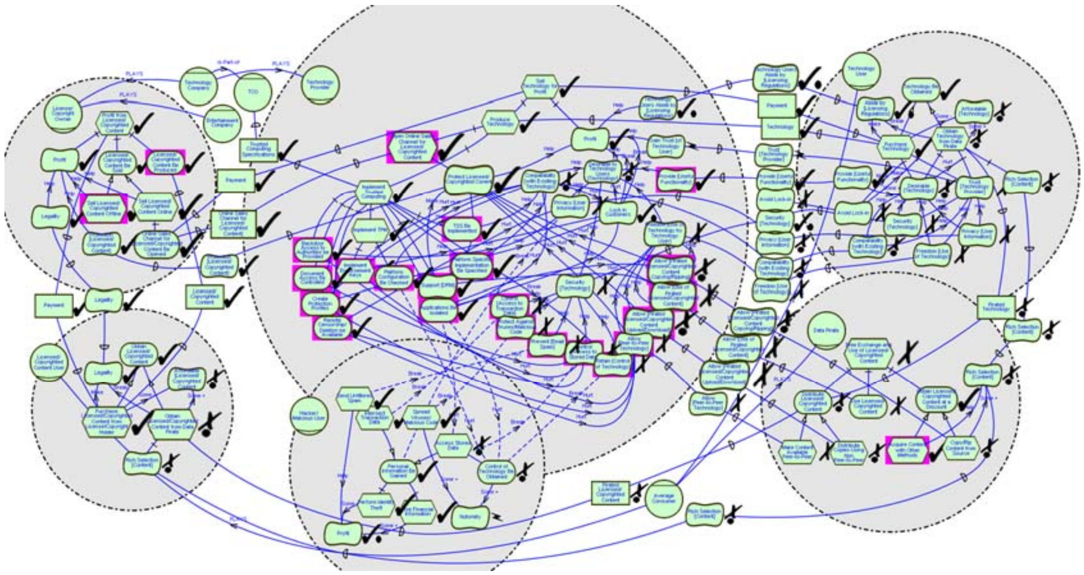

Figure 3. High-Level view of the Trusted Computing Opponent Model

Lessons Learned: i* modeling and evaluation were helpful in describing opposing and complex viewpoints. The evaluation procedure especially helped to provoke changes in the model which, in the opinion of the modelers, improved the quality of the model and forced the modeler to learn more about the domain.

Social Service Case Study

The next study applied agent goal modeling and analysis to a domain involving a real enterprise, including access to a variety of stakeholders. This case study, used in our examples, applied the evaluation procedure in several stages of the project.

Stage 1: The first stage of the project was described in Section **Error! Reference source not found.** as an illustrative application of our example methodology. Here the procedure was applied manually to large models (the largest had 353 elements) in order to analyze and compare the effectiveness of technology options for providing counseling over the internet. The models were created based on transcripts of interviews with several roles in the organization. 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. The organization opted to continue to use a modified version of the moderated bulletin board option already in place, due partially to a lack of resources available to handle online counseling traffic. Results of this study were used to study the use of viewpoints in conceptual modeling (Easterbrook et al., 2005).

Lessons Learned (Stage 1): Although the process of modeling and analysis helped the analysts understand the organization and evaluate technology options, the models created were large and difficult to modify. As this was one of the first large, "real-life" agent-goal model case studies the investigators had undertaken, our initial approach was to model everything. Although our general mandate was to explore new technologies, we did not agree on a common scope for our analysis, and therefore tried to cover everything we learned about the organization in our models. The resulting models were difficult to understand, change, evaluate and were nearly impossible to validate with the stakeholders. We compromised the validation process by picking out smaller model excerpts representing key points in our findings, and presenting these to a group of stakeholders.

Stage 2: 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. One large model focusing on online counseling was created. Again, the model was too large to validate with stakeholders, instead each option was presented, listing the important goals positively or negatively affected by the option in tabular form. This helped us work with the stakeholders to prioritize and select features for an online counseling system. The final outcome was a requirements specification document provided to the organization. Undergraduate students were employed to modify open-source software to meet the organization's specifications. Due to resource limitations and the risks involved in deploying the new system, the organization opted to modify their existing system.

Lessons Learned (Stage 2): Although the scope of this stage of the project was smaller than previous stages, the resulting model was still large and difficult to manage. Evaluation was applied extensively, but, as the models were created using Microsoft Visio, evaluation was done by hand. Manual evaluation over large models was time-consuming and error prone. The trade-off between completeness and model utility came to the forefront, emphasizing the importance of scoping.

Stage 3: A later stage of the project focused on applying enterprise modeling to analyze the knowledge management needs of the organization (Strohmaier et al., 2007). Models were again based on stakeholder interviews, but this time we created a first draft of each model "on-the-fly", with one of the analysts making a model of the interview content during the process of the interview. Models were later expanded, edited and reorganized based on interview transcripts and scoping decisions. The evaluation procedure was applied manually to the resulting 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 reports and presentations, receiving positive stakeholder feedback. The largest model evaluated in the study contained 544 elements, helping to demonstrate the scalability of the procedure. Selected results are reported in Strohmaier et al. (2008).
Our observations in this and other applications of the procedure attest to the model iteration provoked by evaluation. For example, in Stage 3 of this study, a model focusing on communication contained 181 links and 166 elements before evaluation, 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 stage of the study were created by three individuals, with evaluation performed by two individuals, indicating that this effect is not specific to a particular modeler or evaluator.

Lessons Learned (Stage 3): Although the models in this stage were still large, we were much more rigorous with scoping decisions. Each model focused on one specific issue at the core of the organization's Knowledge Management issues. As a result, the models were easier to understand, modify and evaluate. Creating the models "on-the-fly" seemed to make the modeling process easier for the analysts; however, the models were not created in such a way as the stakeholders could view or participate in the modeling process, so this technique did not assist in model validation. Simplified versions of the models and analysis were shown to stakeholders in presentations, with positive feedback. We provided the models and analysis results to the organization, supplemented by textual description. However, it is difficult to know if the stakeholders would have been able to read and get value out of the models and analysis without our guidance.

Although the evaluation procedure prompted us to iterate over the models, our contact with the stakeholders was not frequent enough to allow us to go back to the stakeholders with questions prompted by model changes. Future studies must test for this phenomenon.

Overall Lessons Learned: Applying agent-goal model and interactive, qualitative analysis in the case studies demonstrated the ability of the approach to aid in domain understanding, analysis, decision-making, and communication. Results of our analysis revealed interesting points, and were well-received by participants. However, we found difficulties in the scalability of the modeling process, the scalability of manual analysis and the validation of models and analysis with stakeholders. Ideally, future studies would involve the stakeholder directly in the modeling and analysis process.

Prompted by our case study experience, an exploratory experiment was developed and carried out an in order to test some of the perceived benefits of the procedure.

EXPERIMENT DESIGN AND RESULTS

Observations in case studies have shown that the evaluation procedure described in this work aids in finding non-obvious answers to analysis questions (covering General Desired Characteristics G1 to G4), prompts improvements in the model (E3), leads to further elicitation (E4), and leads to a better understanding of the domain (E4). Our experiment begins to test whether these effects are specific to our procedure or are a product of any detailed examination of a model. We are also interested in how modeling and evaluation experience as well as an evaluators' role in creating a model affect our results.

Design

The experiment materials were taken from a study applying goal-oriented analysis to the sustainability issues for the ICSE conference (Cabot et al., 2009). The study produced a series of models focusing on various actors in the domain of conference planning. The participants of this study, including one of the authors, having knowledge of the domain, were asked to participate in a further study testing the effects of the procedure described in this work. The participants 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.

Models and Questions: Three models (M1, M2, and M3), each created by one or more participants, were selected for use in this experiment. M1 focused on the Publicity and Program Committee Chairs and contained 55 elements, 82 links and 8 actors, M2 focused on the Conference Experience Chair and contained 36 elements, 50 links and 5 actors, while M3, containing 78 elements, 130 links and 15 actors, focused on the General Chair.

Participants were given two different questions (Q1 and Q2) specific to each of the three models. Questions were developed to be non-trivial, and either explored the effects of a particular set of options on high-level goals or asked more general questions related to the possibility of goal satisfaction. For example, Q1 over M1 "If the Publicity Chair distributes materials online and the PC Chair prepares only online proceedings and has only online submissions, how will this affect the significant goals of the actors (acceptance rate, quality of program, diffusion, etc.)?" and Q1 for M2 "If every task of the Sustainability Chair and Local Chair is performed, will goals related to sustainability be sufficiently satisfied?".

Participants: The participants were all current or former researchers in Computer Science or Information Systems disciplines. Three of the participants (P1, P2, and P4) were new to i* before creating the models in question, another participant, P3, had extensive experience with goal models, and one participant (the paper author, P5) could be considered an "i* expert". P1 to P4 were not aware of the specific hypothesis of this study.

Experimental Steps: Due to the small number of participants, we did not split the participants into groups using, and not using, the procedure. Instead participants evaluated models not using, and then using the procedure, examining the changes and additions between results. In this set-up participants would already be familiar with the analysis results before applying the procedure. Therefore, the focus will be on examining the effects of the procedure beyond what can be gained by ad-hoc analysis.

For each of the three models, participants were asked to describe their role in creating the model, record their answers to the analysis questions, and record any question derived from their analysis. They were then asked to familiarize themselves with the evaluation procedure by reading a manual containing an explanation of the interactive evaluation procedure. Participants then re-evaluated each model using the procedure, again capturing questions, model changes and analyzing the differences between their analysis results. They were asked to record the time taken for each step.

After all steps, participants were then asked to answer several follow-up questions: Did model changes improve the quality of the model? 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?

Results and Analysis

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 (G1 to G4). Here, we capture whether or not specific types of changes occurred in the analysis results after applying the procedure, and not a count of these changes. For each model, there are two questions over which results could change with application of the procedure, making six possible areas of change. For example, NP_Q1 to P_Q1 for M1 explores the differences between the answers for Q1 concerning M1 with no procedure (NP) and the procedure (P). For each area, we categorized changes under one or more categories: no change, one or more changes in strength (partial to full satisfaction/denial or vice versa), one or more changes in polarity (a change between one of partial/full satisfied, partial/full denied, and conflict), more elements evaluated (more elements included in evaluation results), and less elements evaluated (elements which were included in evaluation results not included).

Table 4 shows each of the areas of change and how many participants made a particular type of change in that area. For example, in applying the evaluation procedure to Q1 for M2, one participant made no change from their previous results, three participants made changes in strength, and one participant evaluated fewer elements. We exclude data which breaks the changes down by participant, noting only that all participants made changes to their analysis results, with each participant making between 7 and 11 total changes in all areas.

Table 4. Analysis Changes – Participant Sum: Area of Change vs. Type of Change

	M1		M2		M3		
	NP_Q1 to P_Q1	NP_Q2 to P_Q2	NP_Q1 to P_Q1	NP_Q2 to P_Q2	NP_Q1 to P_Q1	NP_Q2 to P_Q2	Sum
No Change	1		1	1	3		6
Change in strength	2	1	3	2		2	10
Change in polarity	3	1		3	1	2	10
More elements evaluated	3	3			1	2	9
Less elements evaluated	1	2	1			1	5
Sum	10	7	5	6	5	7	40

Generally we can see that participants did make changes in their analysis results when applying the evaluation procedure, helping to demonstrate that the procedure helps find non-obvious results. Having a more consistent way to propagate evidence, users make changes in label strength, changes in polarity, notice new paths of elements missed, and notice paths which were not actually affected by the alternative.

Although we see that changes are made, we must question whether these changes produce more accurate results. One participant found the evaluation procedure to be too conservative, marking elements as partially satisfied /conflict that were previously judged to be satisfied/partially satisfied. In such cases, we would hope that the evaluator would use this as a catalyst to modify the model, but in this particular case, the only changes made were before the evaluation procedure was applied. However, the same participants stated for a different model that: "…the evaluation showed the model's weaknesses more clearly." Overall, results seem to reveal inconsistencies between the model and the user's perception of the domain.

Next, we count the changes made to the models not using and using the procedure (E3).

We classify changes to models in several categories. Unlike the changes to analysis results, these results are counted on an individual basis. Table 5 shows the count of each type of change made for each model during each question analysis, where the numbers are summed over all participants. More detailed results showing the breakdown per participant are omitted; however these results show that all participants made changes, with the number of changes made by participants over all models ranging from 6 to 25.

Table 5. Model Changes Summed over Participants

	M1				M2				M3				
Categories	NP_ Q1	NP_ Q2	P_ Q1	P_ Q2	NP_ Q1	NP_ Q2	P_ Q1	P_ Q2	NP_ Q1	NP_ Q2	P_ Q1	P_ Q2	Sum
Add Link	0	0	1	0	5	4	0	2	2	6	0	2	22
Remove Link	1	0	0	0	5	7	3	0	0	9	0	3	28
Add Element	0	0	1	0	1	2	0	0	0	1	0	1	6
Remove Element	1	0	0	0	1	2	2	0	0	3	0	3	12
Change Link	4	3	1	2	0	0	2	4	1	0	0	0	17
Change Element Type	1	0	0	0	1	0	1	0	0	0	0	0	3
Remove Actor	0	0	0	0	0	1	0	0	0	0	0	0	1
Move Link	2	0	2	1	2	3	4	4	0	2	0	1	21

Move Element in Actor	0	0	0	0	1	0	0	0	0	0	0	0	1
Sum	9	3	5	3	16	19	12	10	3	21	0	10	111

Results show that more changes were made during the initial analysis without the procedure (71 changes) than with the procedure (40 changes). These results are somewhat surprising, indicating 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. This leads to a further hypothesis left for future testing: automated procedures would be less likely to provoke model iteration. We note that the participants found 40 additional changes using the procedure to answer the questions for the second time, future studies should make use of two participant groups in order to measure if second round ad-hoc analysis would also produce additional changes. In examining the model quality improvement (E3), 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 look at the number of questions the participants came up with when finding answers to the models (E4). The questions collected are categorized in three ways: a specific question concerning the domain, a question which points out a flaw in the model, and a general comment concerning the need to expand the model. Table 6 shows the results summed over all participants. All participants came up with various types of questions, with the number of questions per participant ranging from 5 to 16. We can see that many of the "questions" were actually participants pointing out flaws in the model, providing further evidence supporting the ability of the procedure to provoke model changes.

Table 6. Questions Found Summed over Participants

	M1				**M2**				**M3**				
	NP_Q1	NP_Q2	P_Q1	P_Q2	NP_Q1	NP_Q2	P_Q1	P_Q2	NP_Q1	NP_Q2	P_Q1	P_Q2	Sum
Specific Question	4	1	4	5		3	1	1	1		1	1	22
Flaw in Model	2	5	2		2	2		1	2	2		1	19
Need for expansion					1	1	2						4
Sum	6	6	6	5	3	6	3	2	3	2	1	2	45

The results in Table 6 show that 26 questions were derived without using the procedure, with an additional 19 derived using the procedure. These results are again interesting in that they show careful examination of a model through ad-hoc analysis leads to further elicitation. We could hypothesize again that this effect may not occur with automated analysis. It is promising to note that even though many questions were found without the procedure, application of the procedure provoked a significant number of further questions, even though the same analysis questions were being evaluated. Future studies need to test whether further ad-hoc analysis would produce the same results.

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 (E4). Specific comments include:

"The procedures helped to identify where there were conflicts (which often indicated problems in the models than a true conflict in the situation), which I did not see just by evaluating "intuitively". When the evaluation procedures resulted in "undecided" labels, it emphasized the problem (or lack of information) in the analysis questions themselves rather than in the models."

"Your automated procedure was overall more helpful mostly because it kind-of offered me somebody to argue with about my own intuitions."

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. In terms of ownership, P1 had a role in creating M1, P2 and P4 had a role in creating M2, and P5 had a role in creating M3. Results do not clearly indicate if the role in creation affects the results measured. The same holds for experience, P5 (the author) who had the most experience with the procedure, did not produce results which stood out significantly from the other participants. Finally, all five participants said they would use the procedure again if they had to evaluate another i* model.

DISCUSSION

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 (E3 and E4). The participants have reported that the procedure provides a better understanding of the model and domain. The experimental design has implicitly tested the simplicity of the procedure (E5), as the participants were able to learn and apply the procedure manually. 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. Experiments could be designed to test the ability of stakeholders to validate and trust analysis results (E6 and E7). Further studies can attempt to determine whether evaluation benefits are specific to the qualitative, interactive procedure introduced in this work, or apply more generally to other agent-goal model evaluation procedures.

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. Case study experience has shown that 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 stakeholders to apply the procedure on their own. The implementation of the procedure in OpenOME should better enable future case study application.

We can use the case studies and experiment presented in the previous sections to assess the effectiveness of the procedure presented in this work in light of the desired characteristics listed in Table 1. We insert qualitative evaluation labels into this table to summarize our assessments (Table 7). The procedure is designed specifically to assess the achievement of objectives using a qualitative scale (G1). The procedure has been proven effective in evaluating and comparing alternatives in multiple case studies (G2). The inclusion of softgoal tradeoffs, contribution links, and propagation of values over such links has allowed the procedure to show how alternatives tradeoff between qualitative goals (G3). Although some problems evaluating complex models were encountered, the implementation of the procedure should make evaluation in these cases less difficult (G5). Evaluation results between different evaluators are consistent, assuming the evaluators make the same human judgments. It is likely, however, that different judgments would be made. This actually provides analysts an opportunity to compare and discuss their decisions, assessing the differences in their underlying assumptions.

We attempt to account for the incompleteness and informality (E1 and E2) of agent-goal models by using a qualitative scale with human intervention. Case study experience has shown that this type of analysis is appropriate for early analysis models. Experience in case studies and the initial experiment have attested to the simplicity of the procedure (E5) and the ability of the procedure to facilitate iteration (E3). The questions the participants produced when evaluating the model in the experiment demonstrate the procedure's ability to facilitate elicitation (E4), although future studies should test this phenomenon in more realistic settings. We have not extensively tested ability of stakeholders to validate analysis results (E6), or the trust stakeholders have in results, although the stakeholders in our social service case study trusted the analysis results presented to them in a non-model format (E7). Future participatory studies should better cover these characteristics.

Table 7. Assessment of Desired Characteristics for Agent-Goal Model Evaluation

	General Characteristics			Early Analysis Characteristics	
G1	Assess achievement of objectives	✓	**E1**	Account for incompleteness	✓.
G2	Explore alternatives	✓	**E2**	Account for informality	✓.
G3	Assist in making tradeoffs	✓	**E3**	Facilitate iteration	✓
G4	Evaluate over complex models	✓.	**E4**	Facilitate elicitation and exploration	✓
G5	Consistent analysis	✓.	**E5**	Transparent and simple	✓.
			E6	Amenable to stakeholder validation	?
			E7	Stakeholders trust results	?

RELATED WORK

Goal concepts are prominent in a number of modeling frameworks, notably in "goal-oriented" requirements engineering (e.g. van Lamsweerde, 2001; Kavakli & Loucopoulos, 2004) as well as in enterprise modeling (e.g. Stirna & Persson, 2007; Rolland & Prakash, 2000). 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 Giorgini et al. (2004), Amyot et al. (in press), Letier & van Lamsweerde (2004) and

Franch (2006). 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. Once the number of alternatives has been narrowed using interactive, qualitative evaluation, more detailed information can be added to the model and 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 (Chung et al., 2000). Previous work has used this procedure evaluate i* models, (see Liu & Yu (2004) for example), assuming that the NFR procedure could be easily extended for use with i*, without describing the necessary extensions, modifications, or additional benefits. We build upon this earlier procedure by introducing aspects which cover agent-oriented concepts, providing steps for application, relaxing 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 model (Maiden et al., 2004). The Process Reengineering i* Method (PRiM) builds on this approach, constructing i* models to understand and redesign business processes and associated information systems (Grau et al., 2008). The methodology introduced in this work is more general, applicable to modeling aspects of an enterprise which may or may not be specific to an information system or process.

Specific modeling processes have also been introduced for enterprise modeling, for example the EKD (Enterprise Knowledge Development) Modeling process described by Stirna & Persson (2007). In this process, goal models are created along with five other types of sub-models, questions are used to drive the creation of inter-model links. Participatory, "consensus-driven" modeling is favored over "consultative" participation. Our agent-goal modeling and evaluation guidelines are similar, in that they try to accommodate the participation of stakeholders or models created solely by analysts. Our method differs in that it focuses on the creation and analysis of one type of model. Future work can investigate the modeling and evaluation process to include complementary model types, as is done by Maiden et al. (2004).

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. Guidelines describing how to use this procedure in the process of enterprise modeling have been presented. We have explored the benefits of the evaluation procedure, including analysis, model iteration, and elicitation by describing application to several case studies and by describing the results of an exploratory experiment

Future work could address the creation of a more detailed agent-goal modeling and analysis methodology which varies depending on characteristics of the domain or modelers and which makes use of other types of complimentary models. The procedure introduced in this work can be expanded in several ways, for example: capturing the rationale and assumptions behind human judgments, expanding analysis in a top-down direction as explored by Horkoff & Yu (2008), and giving users selection over different qualitative scales. The implementation of the procedure in the OpenOME tool is currently being expanded to facilitate some of these features.

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

Amyot, D., Ghanavati, S., Horkoff, J., Mussbacher, G., Peyton, L., & Yu, E. (in press). Evaluating Goal Models within the Goal-oriented Requirement Language. *International Journal of Intelligent Systems (IJIS).*

Anderson, R. (2001). Why Information Security is Hard-An Economic Perspective. In *Proceedings of the 17th Annual Computer Security Applications Conference* (*ACSAC'01).* (pp 358- 365). Washington, DC: IEEE Computer Society.

Cabot, J. Easterbrook, S., Horkoff, Mazon, J., Lessard, L., & Liaskos, S. (2009). Integrating Sustainability in Decision-Making Processes: A Modelling Strategy. In *31st International Conference on Software Engineering – Companion Volume: New Ideas and Emerging Results (NIER'09)* (pp. 207-210). New York: IEEE Press.

Chung, L., Nixon, B.A., Yu, E., & Mylopoulos, J. (2000). *Non-Functional Requirements in Software Engineering*. Norwell, MA: Kluwer Academic Publishers.

Easterbrook, S. M., Yu, E., Aranda, J., Fan, Y., Horkoff, J., Leica, M., & Qadir, R. A. (2005). Do Viewpoints Lead to Better Conceptual Models? An Exploratory Case Study. In *13th IEEE International Requirements Engineering Conference (RE'05)* (pp. 199-208). New York: IEEE Press.

Franch, X. (2006). On the Quantitative Analysis of Agent-Oriented Models. In E. Dubois, & k. Pohl (Eds.), *18th International Conference on Advanced Information Systems Engineering (CAiSE'06): LNCS, vol. 4001* (pp. 495-509). Heidelberg: Springer.

Giorgini, P., Mylopoulos, J., & Sebastiani, R. (2004). Simple and Minimum-Cost Satisfiability for Goal Models. In A. Persson, & J. Stirna (Eds.), *16th International Conference on Advanced Information Systems Engineering (CAiSE'04)*: *LNCS, vol. 3084* (pp. 20-35). Heidelberg: Springer.

Gordijn, J., Petit, M., Wieringa, R. (2006). Understanding Business Strategies of Networked Value Constellations Using Goal- and Value Modeling. In: *14th IEEE International Conference on Requirements Engineering (RE'06)* (pp. 129-138). New York: IEEE Press.

Grau, G., Franch, X., Maiden, N.A.M. (2008). PRiM: an i*-based process reengineering method for information systems specification. *Information and Soft. Tech. 50(1-2), 76-100.*

Horkoff, J. (2006). An *Evaluation Algorithm for the i* Framework*. Unpublished Master's Thesis, Department of Computer Science, University of Toronto. http://www.cs.utoronto.ca/~jenhork/MScThesis/Thesis.pdf

Horkoff, J., & Yu E. (2009). A Qualitative, Interactive Evaluation Procedure for Goal- and Agent-Oriented Models. In *21st International Conference on Advanced Information Systems Engineering Forum (CAiSE Forum'09)*. CEUR Workshop Proceedings.

Horkoff, J., & Yu, E. (2008). Qualitative, Interactive, Backward Analysis of i* Models. In *3rd International i* Workshop* (pp. 43-46). CEUR Workshop Proceedings.

Horkoff, J., & Yu, E. (2009). Evaluating Goal Achievement in Enterprise Modeling – An Interactive Procedure and Experiences. In A. Persson, & J. Stirna (Eds.), *2nd IFIP WG 8.1 Working Conference on The Practice of Enterprise Modeling (PoEM'09): LNBIP, vol. 39* (pp. 145-160). Heidelberg: Springer.

Horkoff, J., Yu, E., & Liu, L. (2006). Analyzing Trust in Technology Strategies. In *International Conference on Privacy, Security and Trust (PST 2006)* (pp. 21-32).

Kavakli, E., & Loucopoulos P. (2004). Goal Driven Requirements Engineering: Analysis and Critique of Current Methods. In J. Krogstie, T. Halpin, & K. Siau (Eds.), Information Modeling Methods and Methodologies (pp 102-124). Idea Group.

Letier, E., & Lamsweerde, A. van (2004). Reasoning about Partial Goal Satisfaction for Requirements and Design Engineering, In *12th ACM International Symposium on the Foundations of Software Engineering (FSE'04)* (pp. 53-62). New York: ACM.

Liu, L., & Yu, E. (2004). Designing Information Systems in Social Context: A Goal and Scenario Modelling Approach. *Information Systems. 29(2), 187-203.*

Maiden, N. A. M., Jones, S. V., Manning, S. Greenwood, J., & Renou, L. (2004). Model-Driven Requirements Engineering: Synchronising Models in an Air Traffic Management Case Study. In A. Persson & J. Stirna (Eds.), *16st International Conference on Advanced Information Systems Engineering Forum (CAiSE'04): LNCS, vol. 3084* (pp. 367-383). Heidelberg: Springer.

OpenOME (2009). OpenOME, an open-source requirements engineering tool. Retrieved January 3, 2010, from https://se.cs.toronto.edu/trac/ome/wiki

Osterwalder, A. (2004). *The Business Model Ontology - a proposition in a design science approach*. Unpublished PhD Dissertation, University of Lausanne, Switzerland.

Rolland, C., & Prakash, N. (2000). Bridging the Gap Between Organisational Needs and ERP Functionality. *Requirements Engineering. 5(3), 180-193.*

Spiekermann, S., Dickinson, I., Günther, O., & Reynolds, D. (2003). User Agents in E-commerce Environments: Industry vs. Consumer Perspectives on Data Exchange. In *15th International Conference on Advanced Information Systems Engineering (CAiSE'03): LNCS, vol. 2681* (pp. 1029). Heidelberg: Springer.

Stirna, J., & Persson, A. (2007). Ten Years Plus with EKD: Reflections from Using an Enterprise Modeling Method in Practice. In B. Pernici, & J. A. Gulla (Eds.), *Proceedings of the 11th International Workshop on Exploring Modeling Methods in Systems Analysis and Design (EMMSAD'07)* (pp 99-108). CEUR Workshop Proceedings.

Strohmaier, M., Yu, E., Horkoff, J., Aranda, J., & Easterbrook, S. (2007). Analyzing Knowledge Transfer Effectiveness - An Agent-Oriented Approach. In *40th Hawaii International Conference on Systems Science (HICSS-40)* (pp 188b). New York: IEEE Press.

Strohmaier, M., Horkoff, J., Yu, E., Aranda, J., & Easterbrook, S. (2008). Can Patterns improve i* Modeling? Two Exploratory Studies. In B. Paech, & C. Rolland, (Eds.), *International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ'08): LNCS, vol. 5025* (pp. 153-167). Heidelberg: Springer.

van Lamsweerde, A. (2001). Goal-Oriented Requirements Engineering: A Guided Tour. In *5th IEEE International Symposium on Requirements Engineering (RE'01)* (invited paper) (pp. 249-263). New York: IEEE Press.

Yu, E. (1997). Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering. In *3rd IEEE International Symposium on Requirements Engineering (RE'97)* (pp. 226-235). New York: IEEE Press.