Requirements Eng
DOI 10.1007/s00766-011-0143-y

ORIGINAL ARTICLE

Comparison and evaluation of goal-oriented satisfaction analysis techniques

Jennifer Horkoff · Eric Yu

Received: 1 April 2011 / Accepted: 12 December 2011

Abstract Goal-oriented requirements engineering (GORE) has been introduced as a means of modeling and understanding the motivations for system requirements. Using models to make goals explicit helps to avoid system failures due to implementing the wrong requirements or ignoring certain stakeholder needs. These models are unique when compared to other models used in system analysis in that their structure naturally lends itself to an analysis of goal satisfaction. Existing work claims that analysis using goal models can facilitate decision making over functional or design alternatives, using criteria in the model. Many different approaches to the analysis of goal-oriented requirements models have been proposed, including several procedures that analyze the satisfaction or denial of goals. These procedures make different choices in their interpretation of the goal model syntax, the methods to resolve conflicting or partial evidence, and in the way they represent satisfaction. This work uses three available tools implementing seven similar goal satisfaction analysis procedures to analyze three sample goal models. Results are reported and compared. The purpose of this comparison is to understand the ways in which procedural design choices affect analysis results, and how differences in analysis results could lead to different recommendations over alternatives in the model. Our comparison shows that different satisfaction analysis techniques for goal models can produce variable results, depending on the structure of the model. Comparison findings lead us to recommend the use of satisfaction analysis techniques for goal models as only heuristics for decision making. Our results emphasize investigation into the benefits of satisfaction analysis beyond decision making, namely improving model quality, increasing domain knowledge, and facilitating communication.

Keywords Goal-oriented requirements engineering · Requirements modeling · Model analysis

1 Introduction

Goal models are diagrammatical depictions of user, system, or stakeholder goals and interrelationships. Goal-oriented requirements engineering (GORE) has been advocated to capture and link technical requirements to social needs, to derive high-level or detailed system requirements using elicited goals, and to capture and compare alternative potential implementations. By focusing on clear motivations for system requirements, GORE techniques aim to ensure the right system is built to address the right problems, helping to avoid costly system failures. For example, GORE techniques include [1–5], with [6, 7], and [8] describing example applications of GORE techniques in practice.

Goal models are distinctive among models typically used in the requirements process, (e.g., UML, BPMN, and DFDs) in that their structure facilitates an analysis of system objectives. As goal models contain links describing the contributing relationships between goals (e.g., help and AND), it is natural to trace these links from the selection of a particular goal to other goals along the path of links,

J. Horkoff (✉)
Department of Computer Science, University of Toronto, Toronto, Canada
e-mail: jenhork@cs.utoronto.ca

E. Yu
Faculty of Information, University of Toronto, Toronto, Canada
e-mail: yu@ischool.utoronto.ca

propagating the "satisfaction" of goals onto other goals [1]. Within the goal model analysis literature, a body of work has focused on applying systematic propagation of goal satisfaction (or conversely, "denial") to their models (e.g., [1, 7, 9, 10]). Such analysis procedures can answer questions like "Will a particular design alternative work in the domain?" and "What are the consequences of its implementation?" Other approaches to the analysis of goal models have been introduced, including application of metrics over model structure [11, 12], application of planning techniques [13, 14], simulations [15, 16], and checks for formal properties [17, 18].

Although several techniques take the same general approach of propagating goal satisfaction throughout a model to answer analysis questions, these approaches differ in several dimensions, including the specifics of propagation through links, interpretation of goal model syntax, measurement choices for goal satisfaction, and the level of participation of the user. It is unclear how these different interpretations and choices would affect analysis results. In this work, we aim to understand the practical consequences of these different procedural choices, how they reflect on the reliability of procedure analysis results, and how they would affect use of evaluation in practice.

Existing work in goal model analysis emphasizes the analytical power of goal model analysis procedures. Such work focuses on the conclusions, which can be drawn from the models, and emphasizing their role as a decision-making tool, helping modelers to choose between alternative system functionality or design configurations. However, previous work by the authors has used literature surveys and experiences from case studies [7, 19] to enumerate benefits of goal model analysis beyond analytical power [20]. For example, such analysis can be used to improve the quality of the model or the understanding of the domain by forcing examination of sections of the model or by checking the contents of the model against user understanding. Careful consideration of the model prompted by analysis or consideration of the analysis results themselves can lead to further requirements elicitation, filling gaps in knowledge. Model analysis can be used as a means of communication between and among stakeholders and analysts concerning the effects of alternatives or properties of the model, aiming for convergent understanding of the domain.

It is difficult to judge the accuracy of analysis performed over high-level, social models capturing the "to-be" space. However, we can begin to judge the reliability of analysis results by comparing results across similar procedures. If results are reliable, similar analysis approaches should produce very similar results over the same models. By performing comparisons to check the reliability of analysis results, we evaluate whether or not certain goal model analysis procedures are best used as a decision-making tool or are better used to achieve other benefits, as described.

In this work, we focus on comparing and analyzing the differences among procedures that propagate satisfaction values forward through model links. To make the comparison, we use existing goal model examples from the literature and apply a selection of available procedures to analyze several alternatives within the example models. We define conventions for comparing differing result formats. Variations in the results are analyzed, including the design alternative each procedure appears to favor. The purpose of the analysis is to understand to what degree variants in procedure design affects analysis results. We use these results to understand potential benefits of goal model analysis and how goal model analysis could be used effectively in practice.

This paper is organized as follows: Sect. 2 provides background on goal modeling; Sect. 3 gives an overview of current approaches to forward satisfaction analysis for goal models; Sect. 4 describes the detailed comparison of satisfaction analysis techniques, including sample models, comparison alternatives, tools, technique details, comparison results, and results analysis; Sect. 5 summarizes goal model analysis techniques beyond satisfaction analysis; Sect. 6 discusses the impact of the results and outlines threats to study validity; and Sect. 6 provides conclusions.

2 Background: goal models

GORE frameworks allow for the representation of one or more goals, which may be derived from the system or system stakeholders, and which may have relationships to other goals, often describing how a goal can be achieved, or if a goal negatively impacts other goals. Such models allow an explicit consideration of system or stakeholder goals in the RE process, allowing analysts to ensure that goals are sufficiently satisfied, and that all proposed features or design alternatives satisfy real needs in the domain. The aim is to improve the likelihood of system success by ensuring that the software (or software changes) plays an effective role as part of a complex socio-technical system. Although goal modeling is not yet widely used in practice, it has been applied successfully in several industrial cases, including air traffic management [6], a not-for-profit organization [7], and health care [8]. Several applications of GORE techniques in practice are summarized in [21, 22]. Example goal modeling frameworks, techniques, or methodologies include KAOS, GBRAM, NFR, i*, Tropos, GRL, and AGORA, described briefly below.

The KAOS methodology introduced a formal goal framework applying AND and OR decompositions between goals describing desired states over entities, achieved by actions assigned to agents [3]. The GBRAM technique guides the elicitation of goals from system activities and artifacts, classifying goals, and associating them with constraints and scenarios [5]. Goals in GBRAM are refined using questions and scenarios, and are represented in tabular form.

The NFR (non-functional requirement) modeling framework aims to represent user intentions in technical systems [1]. The framework uses the concepts of *softgoals,* goals that are not satisfied via clear-cut criteria, AND and OR decompositions among goals, and *contribution links*, representing potentially partial negative and positive contributions to and from such goals. The i* (distributed intentionality) framework [2] incorporates concepts from the NFR framework, including softgoals, AND/OR decompositions, and contribution links, as well as (hard) goals, resources, and dependencies between actors (agents). The i* framework is used as a first stage in Tropos, an agent-oriented system development methodology [4]. A simplified version of i* was used to create GRL (goal-oriented language), which together with use case maps (UCM) constitutes URN (user requirements notation), recently approved as an ITU-T international standard [23]. The Annotated Goal-Oriented Requirements Analysis (AGORA) approach includes goal priorities and methods for solving goal conflicts, selecting alternatives, and measuring the quality of models [24].

For more information concerning existing GORE techniques, the reader is referred to GORE surveys presented as part of [21, 25, 26].

In this work, we focus on systematic analysis procedures propagating satisfaction levels over graphical goal model representations consisting of goals and relationships. We limit our survey to analysis procedures that work over models that minimally support a set of goals linked together by AND/OR and some kind of contribution links. This type of goal model allows analysis of satisfaction using the relationship between goals. We explicitly include contribution links as they are the structure over which analysis interpretations differ the most, likely due to the inclusion of partial or negative relationships between goals. Analysis and propagation through AND/OR links are generally simple Min/Max and not a point of contention in the literature. We focus on analysis procedures that use the structure and the relationships of the model to derive useful information such as the effects of alternative designs or the satisfaction level of critical domain properties such as security. As a consequence of this focus, certain types of models are excluded from our study. For example, models in GBRAM focus on obstacles, scenarios, and decomposition, but do not include partial or negative contributions. Models in AGORA combine AND/OR links and contributions links together in the same structure, making it difficult to apply analysis procedures intended to analyze such constructs separately.

Three example goal models are included in Figs. 1, 2, and 3. These models appear in publications describing goal model satisfaction analysis and are used as examples later in this work. Our choice of these models as examples is discussed in Sect. 4.3. We have recreated versions of each model for inclusion in this work using modeling tools introduced with each work. Figure 1 depicts a Media Shop example in the Tropos framework [9], containing alternatives (linked via OR) related to managing the shop and ordering items. These alternatives are related to softgoals (cloud-shaped) describing the desired non-functional qualities of the system. Figure 2 contains a GRL model of a Wireless Service Provider and Vendor [10]. The system must decide where to place new data and services, using links to softgoals to analyze the effects of alternatives. Figure 3 contains an i* model (from [7]) showing a counseling service for kids and youth, including youth and counselors. Here, the service must decide between different types of online counseling using softgoals in the model.

The reader can note that despite some differences in concepts and notation, the model languages and styles are fairly similar. For instance, each model has goals, although Figs. 2 and 3 further distinguish between goals, softgoals, tasks, and resources. Each model contains AND and OR links, although the visual representation of each differs slightly. All models contain contribution links, although the specific style of the links differs. Figure 1 uses ++/−− for strong and ± for weak positive/negative contributions, respectively. Figure 2 uses ± but with a dot on the top, no dot, or a dot on the bottom, for strong, some and weak contributions. Figure 3 uses words over the links to describe the strength and type of contribution: Make, Help, Break, and Hurt. Figures 2 and 3 contain actors (the large circles) with goals assigned to them within the circles. These models have dependency links showing goal, softgoal, task, or resource dependencies between actors. Despite the differences, the models contain enough conceptual similarities (i.e., AND/OR and contribution links) to make it possible to use these models to compare the results of several satisfaction propagation analysis methods, as is done in Sect. 4.

3 Satisfaction analysis approaches

Having reviewed goal models in the preceding section, we now review analysis approaches that evaluate the satisfaction of goals in such models. Several approaches introduced for the analysis of goal- and agent-oriented models are aimed to determine the satisfaction or achievement of goals. Other methodologies have been

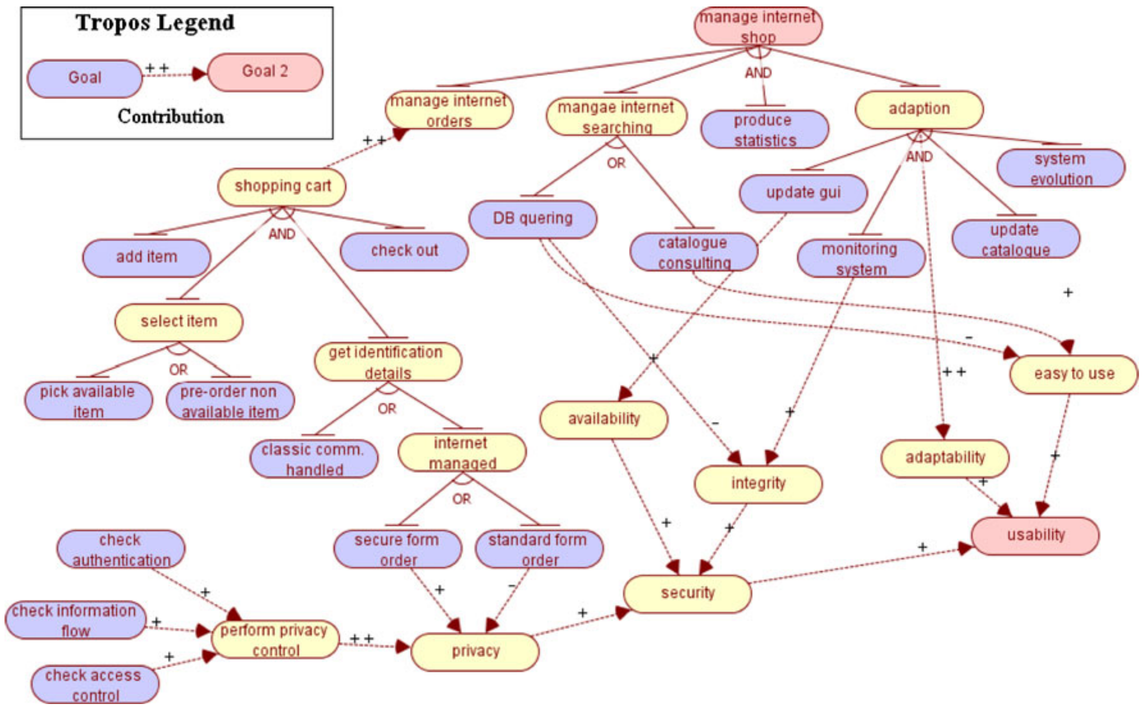


Fig. 1 Tropos actor diagram from the Media Shop example appearing originally in [9]

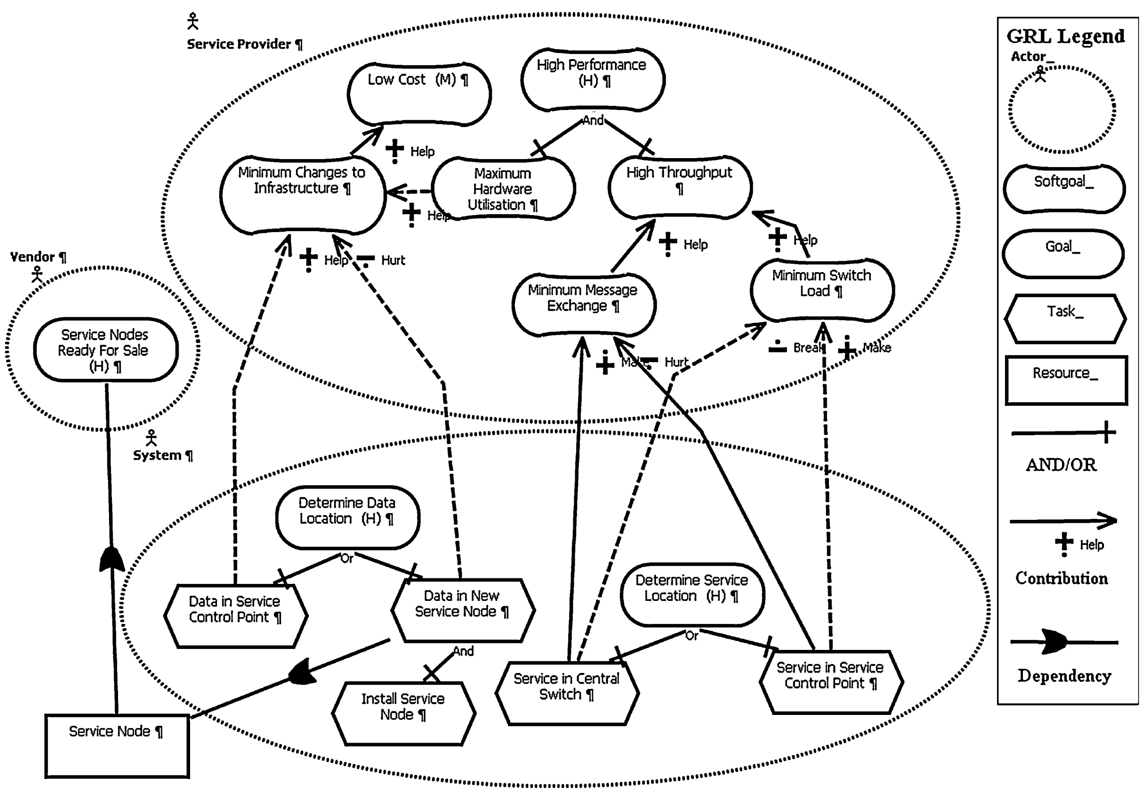


Fig. 2 GRL model of a Wireless Service appearing originally in [10]

developed to measure specific properties such as predictability or risk, or to apply planning, simulation, or model checking to agent-oriented goal models, attempting to find effective system configurations or to detect problems in the high-level system design. As mentioned in the introduction, in this work, we focus on the first type of analysis

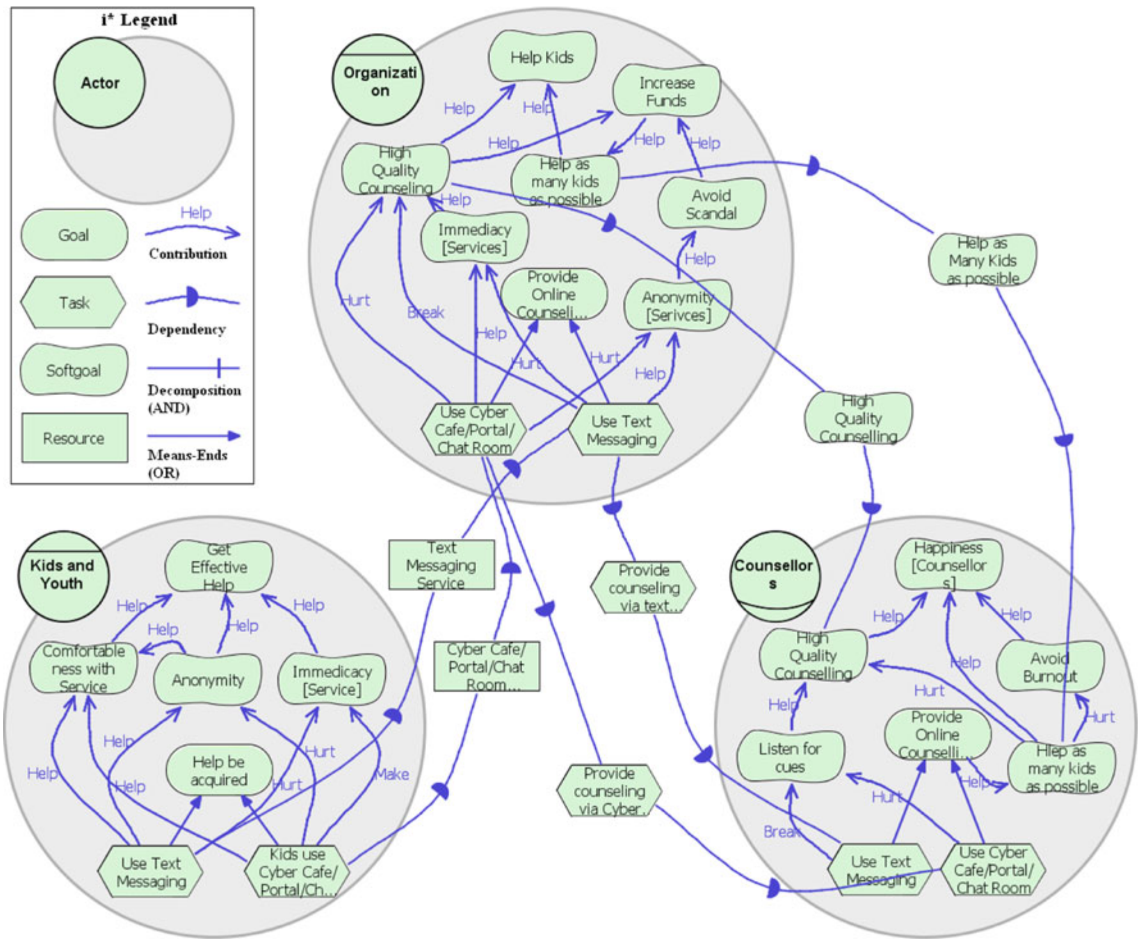


Fig. 3 i* Counseling Service model appearing originally in [7]

approach, measuring goal satisfaction. A brief summary of other GORE analysis approaches are provided in Sect. 5. A more detailed review, including recommendations for selections between different goal analysis categories (metrics, planning, etc.) can be found in [20].

Goal satisfaction procedures start with initial values assigned to the model, reflecting an alternative or question, and then use model links to propagate values either forward (in the direction of the link) [1, 7, 9, 10, 27–29] or backward [10, 27, 30, 31]. These procedures can answer questions like "What is the effect of this alternative?" (forward) or "Can these goals be satisfied?" (backward).

Some satisfaction analysis procedures present results in terms of qualitative labels representing satisfaction or denial, typically using: (sufficiently) satisfied, partially satisfied, (sometimes) conflict, none/unknown, partially denied, and denied, [1, 7, 9, 10]. Several procedures offer quantitative analysis, using numbers to represent the probability of a goal being satisfied or denied [10, 27, 28], or to represent the degree of satisfaction/denial [9]. Other procedures produce binary results, where goals have only one of two values, typically satisfied or not. For example, Maiden et al. analyzes in terms of compliance whether an argument can be made to justify the satisfaction of tasks and resources based on existing requirements [29].

One of the primary distinguishing features between these approaches is their means of resolving multiple incoming values for goals. Goal models often include contribution links representing positive and negative consequences of various degrees (see Figs. 1, 2, and 3 for examples). A goal could receive several different types of contributions at once, positive and/or negative of various strengths. Some procedures deal with such situations by separating negative and positive evidence, making it unnecessary to resolve conflicts [7, 30]. Other procedures make use of predefined qualitative or quantitative rules to combine multiple values [9, 27]. Further procedures are "interactive," using human intervention based on domain knowledge to resolve partial or conflicting evidence [1, 7]. We compare these procedures using example models in Sect. 4, in order to try and determine the practical consequences of these resolutions choices, along with other procedural design choices.

4 Detailed comparison: satisfaction analysis

In this section, we present the details of our comparison, including the selection of procedures for comparison, details concerning these procedures, the selection of example models, model alternatives, and analysis tools. We provide the comparison results, analyzing the potential sources of results variation across models and procedures.

4.1 Analysis procedure selection

We select a subset of the GORE analysis techniques for a more detailed evaluation and comparison. We apply these procedures in our comparison by running them over three alternatives in three models using three tools. Procedure selection is focused on qualitative and quantitative satisfaction analysis techniques that propagate satisfaction levels in the forward direction. Specifically, we select the following procedures: the three GRL procedures (quantitative (GRL-quant), qualitative (GRL-quant), and hybrid (GRL-hybrid)) [10]; qualitative procedure aimed for i* models (i*) [7]; the interactive, qualitative procedure introduced as part of the NFR framework (NFR) [1]; and the forward qualitative and quantitative procedures associated with the Tropos methodology (Tropos-qual), (Tropos-quant) [9]. We provide more detail concerning the specifics of each of these procedures in the next section.

We omit other procedures which propagate forward satisfaction values, such as Maiden et al. [29], AGORA [28], and KAOS [27], as these procedures are too dissimilar to the selected seven to produce a clear comparison. Specifically, [29] differs in the scale of measure and the coverage of propagation. The approach only propagates compliance or non-compliance and not degrees of satisfaction or denial. The procedure propagates compliance originating from only one requirement/task at a time, as opposed to a set of selected initial values as is done in the seven selected procedures. The impact analysis procedure described in [28] for AGORA is focused on change management, detecting conflicts when a new goal is added and analyzing goal achievement when a goal is deleted. When a goal is added, the procedure uses goal characteristics such as security or usability to suggest conflicts between goals. When a goal is deleted, the approach calculates impact on the parent goal using a ratio of the contribution values assigned to the links. Unlike in the selected procedures, this value is not propagated further up the graph. The AGORA procedure can also calculate achieve and obstruct values for the roots goals in the graph. As this type of propagation is very similar to the Tropos quantitative evaluation [9], we omit it from our comparison.

Work in [27] produces degrees of probabilistic satisfaction, but requires additional, specific information in the form of cumulative distribution functions over random variables for goals in the model. The sample models available for each of the other procedures does not contain this information and the information would not be explicitly used by any of the procedures, making a comparison difficult.

We also omit the backward propagation procedures [30, 31] as the form of analysis question between backward and forward propagation is different ("what if?" vs. "is this possible?") making the results difficult to compare.

In this work, we focus on comparing aspects of techniques, which propagates the satisfaction and denial of goals in some way. However, several of these techniques have been expanded to allow further analysis capabilities. The backwards approach in [30] allows for the addition of analysis constraints, conflict restrictions, and finding a minimum cost solution. Asnar and Giorgini [32] expand on [9] to include quantitative analysis of acceptable risk levels and costs. This procedure works over an expansion of the Tropos framework that includes events, risks, and (risk) treatments. Wang et al. [33] adapt the work of Giorgini et al. [30], using goal models to diagnose run-time failures. Amyot et al. [10] use quantitative, qualitative, or hybrid analysis and use per-actor goal priorities added to the models, to calculate an overall numeric satisfaction value for an actor. We do not consider these extended features in this study.

4.2 Selected procedure details

In order to better understand the differences between the analysis procedures and the significance of these differences in terms of results, we describe each procedure in more detail.

4.2.1 NFR evaluation

A qualitative evaluation procedure was introduced as part of the NFR framework with the high-level intention of allowing evaluation of design alternatives with respect to the non-functional requirements of the system, helping to choose the alternative that involves the best trade-offs between system goals. To this end, labels are placed on the graph to indicate the selection of an alternative, these labels are propagated throughout the graph, and the results are analyzed.

The procedure uses the concepts of "satisficed" to represent a sufficient level of goal satisfaction (achievement) and "denied" to represent negative achievement gained through negative contributions. These labels are also used to initiate the procedure by selecting design alternatives. The procedure uses six qualitative labels to represent fully satisficed (√), weakly satisficed (W^+), undetermined (U), conflict (♮), weakly denied (W^-), and denied (X). Weak satisficed/denied refers to the situation where there exists positive/negative evidence toward the satisfaction/denial of a goal, but this evidence is not sufficient to judge the goal as fully satisficed/denied. Undetermined represents the case where no evidence is available. Conflict indicates that an element is both satisfiable and deniable.

In the procedure, the initial labels are propagated from offspring to parent goals using both propagation rules and human judgment. These rules indicate what labels are propagated, given the label of the offspring and the type of link. AND contribution links propagate the minimum value, using the ordering:

$$\natural \leq U \leq X \approx \checkmark$$

while OR links propagate the maximum value. Propagation through other contribution types (+, Some+, ++, −, Some−, −−) is described in Table 1, recreated from Table 3.2 in [1].

The procedure consists of two steps. In step one, all current values are propagated from offspring to parent using the propagation rules. Goals which receive an unknown or conflict label require human intervention. As a softgoal may receive more than one label via more than one contribution link, these labels must be combined into a single label, possibly requiring human judgment. Step two involves the resolution of these softgoal labels. The procedure suggests collecting the labels for one parent node in a bag, allowing for duplicates. In some cases, when the result is clearly satisficed, denied, or conflict, the incoming labels can be combined automatically. In other cases, human intervention is required. The work recommends that all partial values are combined together into one or more full, unknown or conflict labels, and that the final result is the minimum of these combined labels, using the ordering above. If both a satisficed and denied label remain, the result is a conflict. These steps are then repeated until all values have been propagated.

The human judgment required in this procedure is a point of interest. It is up to the analyst to promote or demote values and to try to resolve conflicting values. This process should make use of domain knowledge, including knowledge of the relative importance of each offspring.

Although the rules given describe the promotion or demotion of all partial values, it is mentioned that the procedure can be expanded to allow for partial values as a final label of a node. The last part of the description outlines how the procedure could be modified to allow for weak labels as results, but does not specify the full details of this adjustment. For the purpose of comparison, we follow the original description, forcing all partial values to be promoted via human judgment.

4.2.2 *i* Evaluation*

In [7, 10], and [34], the procedure described with the NFR procedure is expanded to work on i* models, taking into consideration dependency links, and additional intention (goal, task, softgoal, resource) and contribution (some+/some−) types. Specifically, dependency links are treated as a Make (++) link when incoming to a softgoal, and as part of an AND link when incoming to a "hard" (goal, task, resource) intention.

As the description of the procedure in [1] was given only in high-level prose, the work in [7] added details to the procedure, including the treatment for a mixture of links, definition of initial values, propagation from links to links, a consideration of convergence and termination. The conditions for the application of human judgment were relaxed, allowing the user more freedom in their choice. The procedure allows partial values as the end results of evaluation, not encouraging users to promote or demote partial values unless they deem it appropriate as per the domain. As a result, the propagation rules in Table 1 were expanded and modified slightly, as shown in Table 2 from [7].

4.2.3 *Tropos procedures*

Qualitative and quantitative goal model analysis procedures have been introduced and used as part of the Tropos goal modelling approach [9, 30, 35]. These procedures contain similarities to evaluation with NFR models. Work in [9, 35] introduces a procedure that is qualitative and

Table 1 The "individual impact" of an offspring on its parent during the first step of NFR evaluation, recreated from Table 3.2 in [1]

Individual Impact of offspring with label:	Upon parent label, given offspring-parent contribution type:							
	Break (--)	Some-	Hurt (-)	?	Help (+)	Some+	Make (++)	=
X	W^+	W^+	W^+	U	W^-	W^-	X	X
♮	♮	♮	♮	U	♮	♮	♮	♮
U	U	U	U	U	U	U	U	U
√	X	W^-	W^-	U	W^+	W^+	√	√

Table 2 Propagation rules showing resulting labels for contribution links, recreated from [7]

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

propagates evidence forward or "bottom-up," as with the NFR and i* procedures. However, differences exist between the representation of satisfaction and denial and the syntax of the target models.

Goal models targeted by the Giorgini et al. procedure contain events, observable goals that feed values into the goal graph. In early uses of these models [35], there is no explicit use of the idea of softgoals, all goals can take on partial evaluation values. In later uses [9], softgoals are those goals who have incoming contribution links. In this procedure, the degree of satisfaction or denial of goals is represented using predicates over a goal, where multiple predicates can hold at once. The predicates include full evidence of satisfaction, FS, partial evidence of satisfaction, PS, no evidence, N, full evidence of denial, FD, and partial evidence of denial, PD. For example, for goals G1 and G2, PS(G_1), S(G_1), and PD(G_2) may hold at one time. In this procedure, the term "satisfaction" is used to mean that there is at least full evidence that a goal is satisfied. Each goal is assigned two variables, Sat and Den, over the range of {F, P, N}, representing the level of evidence for the satisfaction and denial of a goal, with F, P, and N representing full, partial, and none, respectively. The predicates FS(G), PS(G), PD(G), and FD(G), where G is some goal, are defined as $\text{Sat(G)} \geq \text{F}$, $\text{Sat(G)} \geq \text{P}$, $\text{Den(G)} \geq \text{P}$, and $\text{Den(G)} \geq \text{F}$, respectively.

The separate formalization of positive and negative evidence allows the procedure to be fully automated by a set of propagation axioms that define how predicate values are propagated through links. Conflicts, the presence of both negative and positive evidence, are propagated and not resolved. Human intervention is not used to resolve evaluation values. The propagation rules implemented defining how axioms are transferred from one goal to another emulate the rules described in [1], modified to account for the separation of positive and negative evidence.

Models used with this procedure allow for non-symmetric contribution links. Labels of S or D on a contribution link indicate that only positive or negative evidence is propagated, respectively. An absence of any letter on the link indicates that the values are propagated symmetrically, meaning both positive and negative evidence is propagated.

The qualitative procedure described in [9, 35] is adapted to produce a quantitative version in the same work. In order to propagate quantitative values, the goal model contribution links must be adjusted to contain numerical weights. As with the qualitative version, positive and negative evidence are stored separately. Goals are again given Sat and Den variables, where Sat(G) = c means that there is at least c evidence of Sat(G). Here, the c values range over a numerical interval [inf, sup], where inf represents no evidence and sup represents full evidence. In the examples, a range of [0, 1] is used, both for the satisfaction and denial of goals as well as the weights of contribution links. The rules are adjusted to deal with these numerical values via the introduction of the $\oplus$ operator, used as disjunction or "max," and the operator $\otimes$, used as conjunction or "min." The $\otimes$ operator is defined as typical multiplication. The $\oplus$ operator is defined as follows:

$$x \oplus y = x + y - x \times y$$

In this scheme, the results of contributions indicate the conditional probability of the parent goal being satisfied, given the satisfaction of the child goal. The application of this numerical model to a goal graph creates a Bayesian network, although work in [36] has pointed out flaws in this assumption.

The propagation rules for this method are consistent with those of the qualitative version. In AND links, positive values are combined via conjunction and negative values are combined via disjunction. The reverse holds for OR links. Evidence propagated through partial links (+, −) is combined via conjunction with the numerical strength of the links. Evidence through full links is propagated without change. When combining multiple sources of evidence with the same polarity (all positive/negative) to a single goal, the maximum value is taken. However, the procedure does not promote partial (PS/PD) values to full values, even if multiple sources of evidence are present, making the results not cumulative.

4.2.4 GRL procedures

Several evaluation procedures have been introduced for GRL [10]. One evaluation procedure is purely qualitative (GRL-qual), using the same qualitative scale as [1, 7, 9] but uses a slightly modified set of propagation rules, particularly in the propagation of evidence across dependency links, which are treated like constraints. More specifically, an intention in a dependency link cannot have an evaluation value higher than those of the intentions it depends on. The procedure propagates values in the order of their link types (first decomposition, then contributions, then dependencies).

The most significant difference between this procedure from the procedures in [1, 7] is the avoidance of human intervention via a set of rules that automatically determines the values of softgoals in all cases. The number of each type of qualitative contribution toward a softgoals is counted, and, depending on how these numbers compare to each other, a value is determined. This procedure also differs in its treatment for conflicting values; propagating a "none" value when the number of weakly satisfied and denied values are equal (in the absence of fully satisfied or denied values).

GRL evaluation supports quantitative (GRL-quant) and hybrid (GRL-hybrid) evaluation in addition to qualitative evaluation. In the quantitative procedure, intention satisfaction and denial are represented by a scale from −100 to 100, where −100 represents fully denied and 100 represents fully satisfied. Model contribution links are assigned numerical values expressing their negative or positive contributions within the same −100 to 100 range. The procedure calculates values for softgoals by multiplying the evaluation value by the link contribution strength and then adding and normalizing the values, using a tolerance value to ensure the values are not ±100 unless one of the contributing links and any incoming dependency links are ±100, respectively.

The GRL-hybrid procedure works like the quantitative procedure, but does not expect quantitative values assigned to contribution links. Instead, it take existing qualitative contribution links (make, some+, help, unknown, hurt, some−, break) and converts them to a quantitative number, (100, 75, 25, 0, −25, −75, −100, respectively).

In addition to satisfaction levels, GRL contains the ability to assign qualitative or quantitative importance levels to goals. These values are combined the calculation of values for softgoals. Goal importance levels are also used in the calculation of an overall satisfaction value for an actor. Because not all of the other procedures under comparison support an inclusion of the relative importance of goals, we ignore this feature in our results comparison.

4.2.5 Selected satisfaction analysis techniques: objectives and methods

When defining propagation over goal models, our selected satisfaction analysis techniques make different interpretations of certain concepts and their relationships. These differences can be attributed to different assumptions concerning the use of goal models in practice, including the objectives of goal model application and how goal models would be used as part of a system development methodology.

For example, the NFR framework [1] provided their analysis technique as a means to determine the impact of design decisions on high-level softgoals. Analysis is intended to be applied after iterative stages of elicitation, NFR identification, operationalization, and decision making. The initial definition of the softgoal concept avoided a formal or quantitative definition, in order to allow for user judgment and flexibility in dealing with non-functional requirements.

Similarly, the approach in [7] leaves softgoal resolution to the user in order to allow for an interactive process that compensates for the incompleteness of models in the early RE process. Evaluation is again performed after an iterative process of elicitation and modeling. However, this approach aims to help analysts make decisions over alternatives in the model, as opposed to evaluating decisions currently made. The work encourages modelers to add knowledge gained as part of the evaluation process to the models, in order to improve their quality.

Other procedures (e.g., [9, 10]) apply a more formal definition to the softgoal concept, either using predicate logic or algorithms to determine their satisfaction levels automatically. GRL evaluation [10] acknowledges that goal models can be applied to achieve several purposes, such as assessing goal satisfaction, evaluating design alternatives, deciding on high-level requirements, testing model sanity, and supporting communication. Given the varying purposes for goal model analysis, this work attempts to support a variety of qualitative and/or quantitative analysis approaches. The description focuses on the evaluation algorithms themselves and not how these algorithms fit into an overall modeling and system analysis process.

The Tropos methodology [9] aims to support all software development phases, including early and late requirements analysis as well as architectural and detailed design. Tropos goal model analysis is intended to answer questions such as "given the satisfaction of as set of leaf goals, can root goals be fulfilled?" and "which set of leaf goals (if any) fulfill all root goals?" Presumably, this type of analysis can be applied in both early and late requirements analysis, although the specific role of analysis in these phases is not described.

Overall, by observing the varying approaches to softgoal definition and resolution (interactive/automatic), we see that some approaches treat goal models as an exploratory tool, capturing imprecise and incomplete information, while some use them as more precise definition of system boundaries, intentions, and interactions. The former assumption would assume that user intervention is needed to compensate for model imprecision and incompleteness, while the latter would assume that the model is fit for automated analysis.

Similarly, some procedures allow for the use of available quantitative measures (e.g., [9, 10]), while others avoid such measures (e.g., [1, 7]). These choices reflect an underlying assumption about the potential availability of accurate numerical domain information, which, in turn, reflect assumptions about the types of systems and domains under analysis (accurate metrics or user estimations readily available or not) or the stage of the project when goal model analysis is applied (early, exploratory stages or later requirements, or design stages).

As goal models and goal model analysis procedures can be applied to a variety of domains and can play a role in multiple stages of a project, we make no assumptions about the "right" way to interpret goal model concepts, or the "right" level of assumptions concerning available metrics. The purpose of this exercise is not to find the "best" technique, but to understand to what degree different assumptions about goal concepts and propagation effect procedure results. If the results do not vary significantly, then the choice between available procedures may not be significant and we may assume that satisfaction propagation techniques provide a level of reliability in their analysis results. If, however, results vary widely, then we note that the differing interpretations of goal model concepts are significant and should be used to guide potential users in their procedure selection. Generally, we intend to use this analysis and comparison of results to help understand how and in what contexts these procedures can best be used.

4.3 Selected sample models

As sample models, we select models used by the original authors to introduce the analysis procedures: the Media Shop model [9] (Fig. 1), the Wireless Service Model [10] (Fig. 2), and the Counseling Service model [7] (Fig. 3). We select these models as they are of a sufficient level of complexity to facilitate interesting analysis results, but are large enough to produce results that may be overwhelming. The sizes of the models are in a similar range, 33, 16, and 31 elements for Figs. 1, 2, and 3, respectively. Each of these models makes effective use of the goal model constructs used in each paper. Finally, the models provide three dissimilar domains over which to test the analysis procedures.

4.4 Selected model alternatives

Within each model, we select three alternatives identified in the original papers; selecting alternatives which are most likely to produce the most diverse results when more than three are available. For the Media Shop example in Fig. 1, we select alternatives 1, 2, and 4 from Table 1 in [9]. We modify these alternatives slightly by adding initial satisfied values to the leaf goals not involved in OR relationships; otherwise, the results would not match what appears in [9]. We select these alternatives as they have roughly the most dissimilar initial values, in order to produce a wider range of results. For the alternatives applied to Fig. 2, the Wireless service example, we select alternatives 1, 5, and 6 from Table 7 in [10] for the same reasons. Work in [7] only analyzes one alternative (Use Text Messaging) over the Counseling Service model in Fig. 3. However, as there are two alternative tasks in this model, we select the other alternatives (Use Cyber Café/Portal/Chat Room), and the alternative where both tasks are satisfied, producing three alternatives. The initial values for all 3 alternatives over all three models are summarized in Tables 3, 4, and 5.

4.5 Selected tool support

Each of the seven procedures has provided a tool implementation, with the exception of the NFR procedure. All the tools are freely available for download. We make use of

Table 3 Initial evaluation values for three alternatives used with Figure 1 from [9]

Element	Alt 1	Alt 2	Alt 3 (4)
DB querying	100, FS	100, FS	
Catalog consulting			100, FS
Pick available item	100, FS	100, FS	100, FS
Classic communication handled	100, FS		
Standard form order		100, FS	100, FS
Monitoring system	100, FS	100, FS	100, FS
Produce statistics	100, FS	100, FS	100, FS
System evolution	100, FS	100, FS	100, FS
Add item	100, FS	100, FS	100, FS
Check out	100, FS	100, FS	100, FS
Update catalog	100, FS	100, FS	100, FS
Check authentication	100, FS	100, FS	100, FS
Check information flow	100, FS	100, FS	100, FS
Check access control	100, FS	100, FS	100, FS
Update GUI	100, FS	100, FS	100, FS

Table 4 Initial evaluation values for three alternatives used with Figure 2 from [10]

Actor	Element	Alt 1 (1)	Alt 2 (5)	Alt 3 (6)
Service provider	Maximum hardware utilization	50, PS	50, PS	50, PS
System	Data in service control point	100, FS		
	Service in central switch	100, FS		
	Service in service control point		100, FS	100, FS
	Install service node		100, FS	100, FS
(Dependum)	Service node			100, FS
Vendor	Service nodes ready for sale		−100, FD	100, FS

Table 5 Initial evaluation values for three alternatives used with Figure 3 from [18]

Actor	Element	Alt 1 (1)	Alt 2	Alt 3
Counselors	Use text messaging	100, FS	−100, FD	100, FS
	Use cyber café/ portal/chat room	−100, FD	100, FS	100, FS

each of these tools to apply the procedures to our three example models, redrawing each of the three models in each tool. Specifically, we apply the evaluation techniques using the jUCMNav tool [37] for the three GRL techniques, OpenOME [38] for the i* and NFR technique (with manual adjustments to human judgment criteria in the application of the NFR technique), and the G-R Tool [39] for the two Tropos techniques. A summary of our overall comparison approach including sources, procedures, tools, and models is shown in Fig. 4.

4.6 Conversions, adjustments, and conventions

Although the procedures have much in common, some conversions, adjustments, and conventions need to be made and adopted in order to allow for the results to be more easily compared. We endeavor to only convert formatting, without affecting the results themselves. Cases where adjustments may affect procedure results are changed deliberately in order to test the impact of such changes.

4.6.1 Measurement values

The qualitative techniques use similar but slightly differing labels, for example PS (partially satisfied), ✓ (partially satisfied), and W^+ (weakly satisfied). For this comparison, we will convert all values to a common scale of fully satisfied (FS), partially satisfied (PS), conflict (C), none (N), partially denied (PD), and fully denied (FD). The Tropos procedures produce two results, one for satisfied and one for denied (Sat, Den), while other procedures produce only a single value. We leave the two values as is, without introducing some form of combining automatic values, allowing the reader to make comparisons. The GRL-quant and -hybrid procedures use a scale from −100 to 100, while the Tropos-quant procedure uses 0 to 1.0 for both Sat and Den. We leave these values as is; however, a comparison can be made by dividing the GRL result by 100 and moving it to the Sat (+) or Den (−) side (for example (−37 = 0, 0.37). When selecting initial values to start analysis, we convert FS = +100 = Sat: 1.0 and FD = −100 = Den: 1.0.

When counting differences between qualitative and quantitative results, we use a rough translation of FS/FD = ±95 to 100 (0.95 to 1.0), PS/PD = ±5 to 94 (0.05 to 0.94). We treat N and C as different values, making the distinction between no evidence and conflicting evidence.

4.6.2 Human judgment

Some techniques (i*, NFR) require human intervention to resolve evidence, we indicate these decisions by presenting the results in parenthesis, for example (PS). Whenever possible, the same judgments are used across all alternatives, in other words, the evaluator does not change her mind from one alternative to the next. The judgments made are intended to be reasonable, reflecting the evidence presented by the model.

In applying the NFR procedure, we use the original description, where all final values must be promoted to one of FS, FD, C, or Unknown, and where Conflict or Unknown is selected whenever present in human judgment. The i* approach [7] relaxes these rules for human judgment, allowing the user to choose ignore conflicts or unknowns and decide on partial resulting values.

4.6.3 Cycles

Because the jUCMNav tool did not allow us to draw a two-goal loop, and to simplify the comparison, we remove one of the links in the Counseling model as it appeared in [7]. As a result of this and differing human judgments, results derived may not match results in [7].

4.6.4 Dependency links

As the NFR and Tropos procedures do not explicitly support dependency links, we have treated these links to *make* (++) contribution links when using these procedures over models with dependency links, for example, if x depends on y, then y makes x.

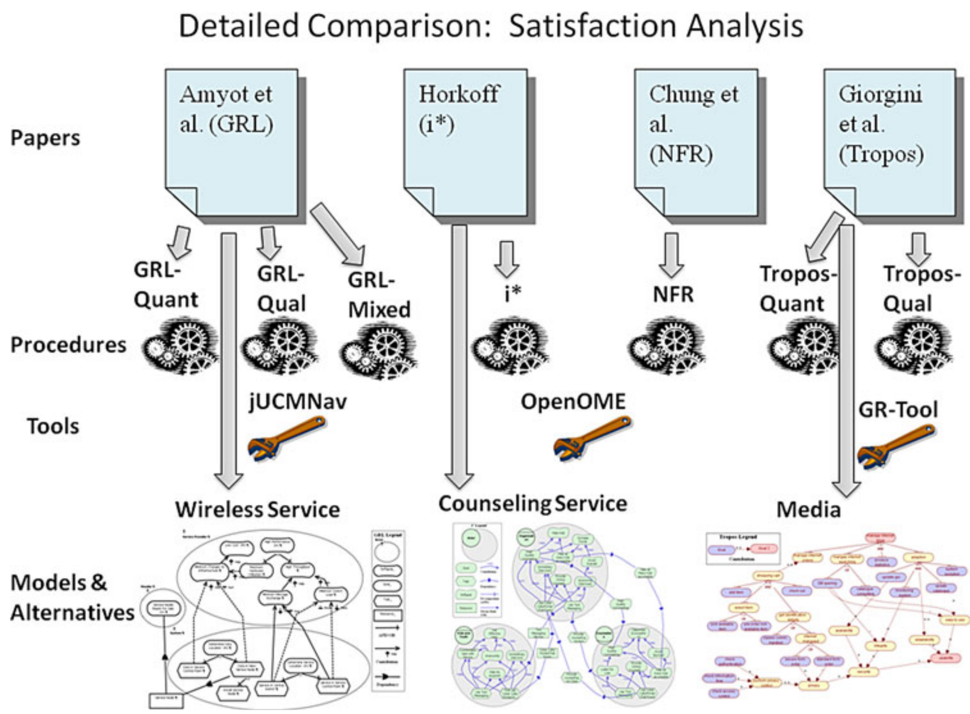


Fig. 4 Summary of the comparison process for the seven procedures over three modeling using three tools

4.6.5 Some+ and some−

The NFR and Tropos procedures do not support the difference between help and some+ (hurt/some-) contribution links. Some+/Help and Some−/Hurt are treated identically in both the i* and the GRL-Qual procedure, only potentially effecting human judgment in i* analysis. In order to equalize the ability of the procedures, Some± links are treated as help and hurt links when undergoing these four evaluation procedures (NFR, Tropos, i*, and GRL-Qual). The GRL quantitative and hybrid procedures automatically convert Help and Some+ links to 25 and 75, respectively, although the quantitative algorithm allows the optional definition of link-specific numeric values. For the application of the GRL-hybrid procedure, we have converted all Some± links to Help/Hurt, respectively, meaning they all have a value of ±25. If we made the same conversion for GRL-Quant, results for the GRL-Quant and GRL-hybrid would be very similar to each other, but not identical, due to algorithm differences [10]. However, in order to understand the significance of numeric label selection on procedure results, we have modified the models to give help/hurt links a value of ±50 in the GRL-Quant algorithm, while the hybrid algorithm retains the original conversion. Quantitative evaluation result will now differ from what appears in [10]. As it is, there are only two Some± links in all of our sample models, both in the Wireless Service Model.

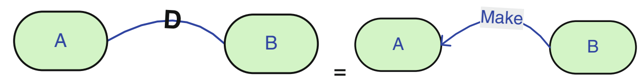
4.6.6 Link symmetry

The Tropos procedures contain contribution links that can be asymmetric, propagating only positive (s) or only negative (d) evidence, while other approaches have symmetric links, propagating both positive and negative evidence. The model and results in [9] are intended to reflect symmetric links; however, a closer observation reveals that the results presented in the paper and the implementation of the G-R Tool only use asymmetric positive links, i.e., only positive evidence is propagated. We use this convention in our trials, partially as a means to determine its impact. Results in this work differ from results in [9] only in value of one goal, integrity, in Alternative 2. This goal has a value of partially negative in our results. We suspect this difference is due to an omission of a hurt link in the sample file available with the GR-Tool when compared to the model as shown in [9], i.e., the results in [9] reflect a model slightly different than the model shown in [9] and in Fig. 1. Conversions, adjustments, and conventions used between procedures are summarized in Table 6.

4.7 Results

We provide tabular analysis results for all three models in Tables 11, 12, and 13 in the Appendix. In these tables, we list the model elements (goals, softgoals, task, resources) on the left, including the actor whose boundary the element appears in. The next set of columns, Alternative 1, presents the results of all seven procedures for the first alternative. Alternatives are distinguished by initial values extracted from the papers, listed in Tables 3, 4, and 5 and marked in the tables below with an asterisk (*). The next two sets of columns, Alternative 2 and 3, provide the same information for the second and third alternatives extracted from the

Table 6 Summary of conversions and conventions used between the seven analysis procedures

	GRL-Qual	i*	NFR	Tropos-Qual	Tropos-Quant	GRL-Quant
Analysis Results	✓ ✓• ⚡ ✗• ✗	✓ ✓• ⚡ ? ✗• ✗	√ W+ U W- X	Sat: S, PS, N Den: D, PD, N	Sat: 0 to 1.0 Den: 0 to 1.0	-100 to 100
Analysis Results Conversion	✓=√ = Sat: S, ✓•=W+=Sat: PS, ?=U, ✗•=W-=Den: PD, ✗=X=Den: D, None=N				Sat_Tropos = +GRL/100 Den_Tropos = -GRL/100	
Conversion for Results Comparison	S = 95 to 100 (Sat: 0.95 to 1.0), D = -95 to -100 (Den: 0.95 to 1.0), C != N		PS = 5 to 94 (Sat: 0.05 to 0.94) PD = -5 to -94 (Den: 0.05 to 0.94)			
Dependency Links	Supported	Supported	Not Supported	Not Supported	Not Supported	Supported
						
Some+ and Some- Links	Supported	Supported	Not Supported	Not Supported	Not Supported	Supported
	Some+ = Help, Some- = Hurt					
Contribution Link Symmetry	Symmetric	Symmetric	Symmetric	Asymmetric Positive	Asymmetric Positive	Symmetric
Human Judgment Initiation	Not Used	Conflicting or partial values	Conflicting or partial values not including ⚡or U	Not Used	Not Used	Not Used
Human Judgment Results	Not Used	Any value	No partial values	Not Used	Not Used	Not Used

paper. Values in parentheses are the result of human judgment. We use the conversions in Table 6 to identify results which differ. When the result from one procedure differs from the rest for one element per alternatives, we highlight this results using bold. When there is a significant difference among the results for two or more procedure, we highlight the whole partial row for that element in that alternative.

4.8 Results analysis

Examining the analysis results over all the models, we can see that each alternative over each model produces differences in results. The differences are often significant. The observed differences include differing values for the top-level goals. For example in Alt 1, Help Kids in Organization is partially satisfied, partially denied and conflicted in the results of three different procedures. We also see that the two types of numeric procedures (GRL and Tropos) differ in how they resolve and combine numbers, sometimes resulting in drastically different results (see for example, Help as Many Kids as Possible in Organization). We analyze the differences first on a more detailed level, with the purpose of understanding why these differences occur. Then, we analyze the differing choices between alternatives that these differences may cause.

4.8.1 Result differences

We count the number of model elements that have differing results for each alternative, with the results summarized in the first column (All) for each alternative in Table 7. For

Table 7 Count of the number of differences in the results over all model elements per alternative

	Alt 1				Alt 2				Alt 3				Totals
	All	Dep	Sym	SG	All	Dep	Sym	SG	All	Dep	Sym	SG	All
Media shop	4	0	1	3	8	0	4	4	5	0	1.5	3.5	17
Wireless service	7	0	0.5	6.5	7	0	0.5	6.5	7	0	0.5	6.5	21
Counseling service	28	7	10.5	10.5	27	7	8	12	27	11	8.5	8	82
Totals	39	7	12	20	42	7	12.5	22.5	39	11	11	18	120

example, in the Alternative 1 columns of Table 12 (Wireless Service Model), seven of the elements have results which differ, where one or more of the results for the seven procedures were bolded in a total of seven rows. This is reflected in the Wireless Service Model row in Table 7.

We can attribute these differences to a combination of analysis procedural choices and structural characteristics of the three models. We can identify three procedural choices that result in the majority of the result differences: (1) the different treatment for dependency values (Dep)—GRL treats them as constraints while the other procedures treat them as requirements; (2) the asymmetry of links (Sym)—the Tropos procedures are not propagating negative values; and (3) different methods for the resolution of values for softgoals (SG)—the NFR procedure insists that final values must not be partial, the NFR and i* procedures allow value promotion, and the GRL does not include the concept of a conflict.

We count the number of differences that are caused by each of these three choices (Dep, Sym, and SG, respectively), displaying these counts in Table 7. In counting the choices, sometimes results for one alternative are caused by more than one procedure choice. In this case, we divide the count by the number of causes, so that the sum of the individual counts adds up to the total number of elements with differences. For example, looking at the seven differences in the first alternative of the Media Shop model, six differences are caused by differences in the way softgoals are resolved, while one difference, for the softgoal adaptability, is caused in part by differences in softgoal resolution (PD vs. FD for i* vs. NFR) and in part by the asymmetry of links, with negative values from easy to use not propagated to this softgoal. We can observe from the sum in the last column of Table 7 that different methods for softgoal resolution (SG) accounts for a slightly larger percentage of the result differences when compared to different methods for dependency propagation (Dep) or link symmetry (Sym).

The number of differences in the results for the first two models is similar, ranging between 4 and 7. These models have a similar structure in that the links are mostly AND/OR links, the models have little or no dependency links, and there are few softgoals when compared to the Counseling model. Statistics concerning the elements and links in each model can be seen in Tables 8 and 9. Presentation of the Media Shop model in [9] distinguishes between soft and hard goals, although presentation in the tool, as reflected in Fig. 1, does not. Generally, in this model, all goals with incoming contribution links are softgoals. We can examine the ratio of Softgoals to goals for each of the three models: 0.27, 2.33, and 6.0, for the Media, Wireless, and Counseling models, respectively. In other words, there

Table 8 Element statistics for each of the three example models

	Goals	Softgoals	Tasks	Resources	Total
Media shop	26	7	0	0	33
Wireless service	3	7	5	1	16
Counseling service	3	18	8	2	31

are roughly 4 goals for every softgoal for the Media model, 2 softgoals for every goal in the Wireless service, and 6 softgoals for every goal in the Counseling model. Because we have identified softgoal resolution (SG) as one of the most significant causes of result differences, models with more softgoals will have a greater divergence. Similarly, we can look at the ratio of contribution to AND/OR links: 0.95, 1.43, and 5.4, respectively. The Media Shop model has a roughly equal number of contribution and AND/OR links, while the Wireless model has about 1.5 times more contribution links, and the Counseling model has more than 5 contribution links for every AND/OR link in the model. The Counseling model also has many more dependencies (12 vs. 2 or 0) when compared to the other models. Similar to the presence of softgoals, our results show that models with many contribution (Sym) or dependency (Dep) links are also more likely to produce different results between procedures.

4.8.2 Alternative selection

Although the examination of differences at a detailed level is interesting, the alternative (solution) that the results would lead evaluators to select is more important. Several of the selected analysis procedures emphasize use of analysis procedures to evaluate the impact of decisions over alternatives [1], select an alternative [7], or evaluate an alternative [9, 10]. The purpose of these activities is to make a selection over one or more high-level design alternative in the model. We use the results of our comparisons of the selected procedures over the example models in order to select one alternative set of design options in each model. These alternatives have been listed in Sect. 4.4.

We rank the alternatives for each model (1st, 2nd, or 3rd), using each of the seven procedures, presented in Table 10. In this table, the rows represent the selection result for each procedure, while the column represents the model and alternative. For example, the third column (Wireless Service, Alt 2), fifth row (GRL-Mixed) shows that after applying the GRL-Mixed procedure to Alternatives 2 in the Wireless Service Model, we would select this alternative as the first choice out of the three alternatives for this model. Similarly, Alt 3 would be the second choice, and Alt 1 the third choice. As there is no systematic way in

Table 9 Link statistics for each of the three example models

	AND	OR	Depend.	Make	Some+	Help	Hurt	Some−	Break	Total
Media shop	12	8	0	3	0	13	3	0	0	39
Wireless service	3	4	2	2	4	1	1	1	1	19
Counseling service	0	6	12	1	0	22	8	0	2	51

any method to select an alternative given the goal evaluation results, we pick the alternative with the most strongly satisfied criteria (softgoals, high-level hard goals). For the quantitative procedures, we decide to select an alternative by adding the results over criteria elements. The sums for the criteria goals are listed in parenthesis after the rank in Table 10. In our example, the sums are 161 for Alt 2, 98 for Alt 3, and 48 for Alt 1. Note that this approach is not recommended by any of the authors in [9] or [10]; however, the presence of numbers makes this approach feasible and tempting.

The table contains several entries with "or" or with "?". In these cases, it is difficult to make a ranking decision over qualitative values for each alternative, which show trade-offs over several goals. For example, Alt 1 and 2 results for the NFR procedure in the Wireless Service Model differ by two goals; either Minimum Message Exchange is FS while Minimum Switch Load is FD, or the inverse. It is difficult to make a decision in this case, without further information, such as the relative priorities of these goals. We have indicated this difficulty in Table 10 by indicating that Alt 1 and 2 for this model (Wireless Service) and procedure (NFR) can be each ranked either first or second. In the Counseling Service model, none of the alternatives satisfied the important goals of each actor sufficiently. As a result, we feel it is not possible to make a selection over the available alternatives, looking at the qualitative results over this model an analyst would likely suggest that none of the alternatives be selected. We indicate this in Table 10 by placing "?" instead of a ranking among alternatives. In these cases, the model could be refined to include further criteria or further alternatives could be suggested.

In the Media Shop model, the selection of alternatives is generally consistent, with the exception of the first GRL-Quant results. In the Wireless Service Model, we see more ranking differences, especially with the GRL-Mixed procedure. These differences demonstrate that differing decisions over procedure conventions have the potential to produce differing alternative selections—even in models with few softgoals, contributions, or dependency links. The selections for the Counseling Service model also differ between procedures, not surprisingly considering the differences in analysis results. In this model, it is often difficult to make decisions over alternatives, especially when analyzing qualitative results.

5 Related work: other approaches to goal model analysis

Work has endeavored to analysis goal models using methods other than propagation of goal satisfaction. Several approaches aim to measure qualities over the domain, such as security, vulnerability, and efficiency, using metrics over constructs in the model (e.g., [11, 40, 41]). These procedures can answer questions like "How secure is the system represented by the model?" or "How risky is a particular alternative for a particular stakeholder?" For example, Franch et al. [41] introduce the means to calculate global or local metrics over i* SD models using classifications and weights of actors and dependencies in an SD model then expand this approach [11] to work over i* Strategic Rationale models, developing a framework which allows for qualitative or quantitative, automated, or interactive metric calculation.

Methods have applied AI-type planning to find satisfactory sequences of actions or design alternatives in goal models (e.g., [42–47]). These procedures can be used to answer questions such as "What actions must be taken to satisfy goals?" or "What are the best plans of action according to certain criteria?" For example, Bryl et al. [42] aim to find satisfactory delegations (assignment of dependencies) in a social network represented via goal model by iteratively finding plans within the model that fully satisfy all actors and then evaluating the plans in terms of cost, similar to the metrics used in [11].

Several approaches have added temporal information to goal models to allow for simulation over the network represented by model constructs (e.g., [46–50]). In these approaches, a particular scenario is simulated, and the results are checked for interesting or unexpected properties. These procedures can answer questions like "What happens when a particular alternatives is selected?"

Several approaches provide ways to perform checks over the models supplemented with additional information (e.g., [44, 50–52]), allowing users to ask questions like "Is it possible to achieve a particular goal?" or "Is the model consistent?" For example, in [51, 52], Fuxman et al. convert i* models to Formal Tropos, supplementing the models with first-order linear-time temporal logic statements to represent desired constraints, and a model checker is used to validate properties and check for consistency.

Table 10 Ranking of alternatives for each model based on analysis results

	Wireless service			Media shop			Counseling service		
	Alt 1	Alt 2	Alt 3	Alt 1	Alt 2	Alt 3	Alt 1	Alt 2	Alt 3
GRL-Quant	2nd (112)	1st (212)	3rd (63)	3rd (224)	2nd (232)	1st (450)	2nd (−296)	1st (−76)	3rd (−375)
GRL-Qual	1st or 2nd	1st or 2nd	3rd	2nd	3rd	1st	?	?	?
GRL-Mixed	3rd (48)	1st (161)	2nd (98)	2nd (225)	3rd (193)	1st (287)	3rd (−249)	1st (50)	2nd (−183)
i*	1st	2nd	3rd	2nd	3rd	1st	2nd	3rd	1st
NFR	1st or 2nd	1st or 2nd	3rd	2nd	3rd	1st	?	?	?
Tropos-Qual	1st	2nd or 3rd	2nd or 3rd	2nd	3rd	1st	?	?	?
Tropos-Quant	1st (3.0, 1.0)	2nd or 3rd (2.625, 1.0)	2nd or 3rd (2.625, 1.0)	2nd (3.25, 1.0)	3rd (3.25, 1.5)	1st (3.75, 0)	3rd (1.25, 2.0)	1st (1.75, 1.5)	2nd (3.0, 3.0)

Although the checks are automatic, an iterative process of manually defining the bounds of the model checker is often required.

6 Discussion

In this section, we discuss the impact of our comparison results, including the benefits of forward satisfaction analysis procedures, and how results can shape the use of such procedures in practice. We consider additional factors that may have further affected comparison results, such as goal priorities, model size, and inconsistent human judgment. Finally, we consider threats to the validity of our comparison study.

6.1 Results discussion

The results of our comparison in Sect. 4 show that the selected procedures produce significantly varying results when analyzing the selected alternatives over the sample models. We see that differing assumptions concerning goal concepts and propagation can have a noticeable effect on analysis results. We believe that the presence of the differing results is significant. Depending on the analysis procedure selected, potential users could make widely different conclusions about the effects of system alternatives on domain goals. These conclusions could lead to differing selections over functional or design alternatives represented in the models.

For example, in Table 10, if results from the analysis procedures were followed without question, the GRL-Mixed analysis procedure would lead users to select Alternative 2, where the service is in the service control point and the data are in a new service node. Results from the i* procedure would lead users to select Alternative 1, where the service is in the central switch and where the new service node is not installed. These are very different decisions using the same criteria in the same model.

Our results have shown that the structure of the underlying model makes a significant difference in the consistency between analysis results. Specifically, the presence of many softgoals, dependencies, and contribution links decreases confidence in analysis results. As softgoals and contribution links are intended to represent "fuzzy," flexible concepts and effects, which are difficult to quantify or formalize, and dependencies represent often represent high-level, social interactions, it is reasonable that different interpretations of these constructs causes differences between analysis results. However, these social and non-functional constructs are particularly useful in early RE analysis, as argued by [2, 4].

We can observe that making decisions over models with many "criteria" goals can be especially difficult, as shown with the high number of "?" in Table 10 for alternative selections in the Counseling Service model. Results in Table 10 show that quantitative values give a more fine-grained evaluation of the results when compared to qualitative procedures, and that this granularity helps to distinguish between alternatives. These results could lead to a recommendation of quantitative procedures such as Tropos-quant or GRL-Quant for decision making between alternatives. However, differences in quantitative interpretations of goal model constructs can produce differing results, as seen in Table 10. There is a danger in placing too much precision in such numbers, which can actually be thought of as finer-grained qualitative estimates. Operations such as addition, although tempting, lead to a false sense of precision. We observe a fundamental trade-off between qualitative and quantitative procedures. Qualitative procedures lack precision and can make selecting an alternative over multiple criteria goals difficult. Quantitative procedures are more precise, which can help to better differentiate between alternatives; however, numbers are affected by procedure design choices, such as conversions from qualitative to quantitative and choices over propagation rules. Furthermore, it is tempting but ill-advised to perform mathematical operations, such as addition, over these results. The precision offered by quantitative procedures needs to be taken with a grain of salt, treating numbers as estimates, with uncertainty increasing with subsequent propagation. The difference between 0.5 and 0.25 may be helpful, but the difference between 0.07 and 0.10 is likely not significant.

The lesson from these observations should not be that goal models should not be analyzed systematically, but that when used in decision making, the analysis process and results should be considered as a heuristic and always be interpreted in the context of the domain. In fact, the process of modeling and evaluating may be as useful as the results, as the process may force the evaluator to examine the model and their domain knowledge and assumptions [7]. Although analysis results for "softer," more social models (softgoals, contributions, dependencies) show a greater variation than "harder" models (AND/OR, hard goals), users should not be discouraged from performing analysis over softer models. In fact, the fuzzy nature of these models calls for systematic analysis in a process of elicitation and specification. However, the role of this analysis should be clear. The variances in the comparison results from Sect. 4 emphasize that the benefits of forward satisfaction techniques may lie less in their ability to facilitate decision making, and more in their ability to provide other benefits, such as improved domain understanding, communication, scoping, and elicitation. For softer, more social models, interactive analysis can encourage a learning process, forcing users to question their assumptions about the model and domain, evolving the model from a draft to a sufficient level of completeness and accuracy. When this process is performed in a group setting, it can facilitate interesting discussions that promote convergent understanding [19].

Evolution over softer models can lead to a more clearly specified, more stable, harder model, over which analysis results can be more reliable. In fact, if there is enough available time, goal model users could apply both interactive qualitative and then automatic quantitative analysis as part of the same process. We can describe such an approach using the concept of early and late RE from the i* and Tropos approaches [2, 4]. These stages could lead into a design of system architecture, as in [4]. See Fig. 5 for an example high-level development process using both styles of analysis.

Both the i* and NFR analysis procedures specifically aim for early RE, supporting high-level understanding and decision making in the absence of specific metrics. The GRL and Tropos approaches leave application open to either early or late RE stages. However, it is difficult to acquire formal or quantified domain information in the early stages of understanding, especially if aiming for completeness. Therefore, these procedures are likely better suited to later stage models, where more detailed information is available. After qualitative, interactive evaluation such as the i* or NFR procedures has been applied, during later stages of requirements analysis, more precise, quantitative forms of evaluation such as the Tropos-quant or GRL-quant can be applied to relatively stable models, producing results that are likely to be more detailed and accurate, and which may be better suited to decision making.

In this study, we were not able to measure the changes prompted by the process of analysis and the consideration of results. Work in [19] uses empirical studies with individual students to attempt to measure model changes and questions prompted by application of the i* procedure described in [7]. This work found that both ad-hoc analysis (manual analysis not using any specific procedure) and the i* procedure produced a small amount of model changes. This result is attributed to the artificial nature of the study, including a lack of motivation for students to improve models over example domain. Future work should test the ability of this and other analysis procedures to improve the quality of models in more realistic settings. These or other studies should apply the analysis procedures in a collaborative setting with groups of stakeholders and/or analysts, such as in [19], to test the ability of each procedure to facilitate useful communication.

Generally, when selecting any type of analysis procedure for goal-oriented models, users should assess the intended purpose of such analysis, for instance: domain understanding, communication, model improvement,

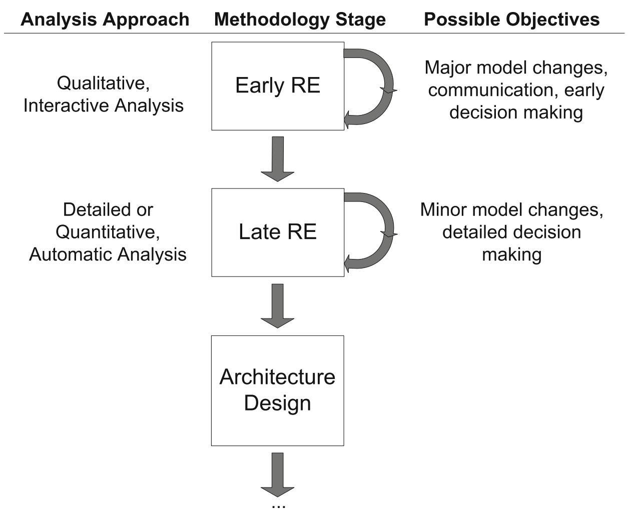


Fig. 5 Example high-level development process using both qualitative and quantitative goal model analysis

scoping, requirement elicitation, requirements improvement, and design. Goal model analysis procedures, as reviewed in Sects. 3 and 5, can be summarized over several dimensions, including the type of analysis questions they can answer, the involvement of users, the notation supported, and the specific information required. Work in [20] suggests a mapping between procedures and benefits based on these characteristics. For example, if model analysis is needed to help make scoping decisions, the mapping recommends trying goal model analysis approaches, which support agent-orientation, allowing users to draw stakeholder and system boundaries. The entire mapping between suggested procedures and benefits can be found in [20].

6.2 Threats to validity

Although the analysis of results over the three sample results has produced useful observations, there are some threats to the validity of our study and comparison. First, the analysis was only performed over three models. Although an effort was made to pick models, which had differing structures covering different domains, further analysis over more sample models are needed to increase confidence in our results.

The models used in our comparison were of small to medium size to facilitate comprehension of results. If larger models were used, the variance in results may be even larger, as larger models mean longer propagation paths away from the common starting point of initial alternative labels. The heuristic nature of forward satisfaction analysis procedures should be especially emphasized for larger models.

Manual classification of the underlying reasons for results differences could be subject to error and interpretation. In fact, differences are often propagated; an initial difference in results for one intention is propagated to produce differences for further intentions. In this case, the cause for the difference was counted once for each intention where it appeared, whether that intention originated the results difference or not. We felt that this convention provided the greatest measure of the impact of individual procedure choices causing results differences.

When running analysis procedures over models, we chose to use the implementation available in associated software tools. The available tools are produced and maintained by research institutions and are not commercialized. It is possible that the implementations may produce erroneous or unexpected results. Whenever possible, our results were checked against results available in the originating papers, with differences analyzed and explained. An example of unexpected results was the presence of only asymmetric propagation in the GR-Tool when the associated paper clearly meant for the symmetry of contribution links to be manually specified and default to symmetric behavior. We could have chosen to manually propagate evidence as specified in the paper and not as implemented in the tool; however, doing so may have introduced errors in either the manual propagation or our interpretation of the procedure, especially for the quantitative propagation. We chose to follow the implemented procedure in all cases.

Comparison of results required an introduction of conversion and conventions as described in Sect. 4.6. It is possible that such conventions may affect the comparison and classification of our results. However, we believe that these effects are minimal. For example, using a convention of PS = 10–90 instead of PS = 5–94 would have a nominal affect on our difference counts.

Certain factors, if included in the design of our comparison, may have had a further impact on results. For example, in our comparison, we have not made use of explicit measures for goal importance available in some of the procedures, e.g., [10]; however, the implicit importance of goals captured in the positioning of the model can be helpful. For example, in Fig. 3, there is obviously a "top" goal for each actor, and these goals could be given more weight when making decisions. Future improvements to these procedures could describe how to make decisions between alternative results.

When applying human judgment in applicable procedures, we chose resulting labels consistent between procedures and consistent with the structure of the model. Work in [19] studied the individual application of interactive goal model analysis procedures by students and found that users often made decisions inconsistent with the structure of the underlying model. For example, a set of incoming labels {PS, PS, PS} may be combined together as a Conflict (C). If users make inconsistent judgments, the variance between the results of interactive and automatic procedures would vary even more significantly, as automatic procedures make automated decisions that reflect their assumptions over the underlying model concepts. Future work in this area could highlight inconsistencies between the model structure and user judgments. The occurrence of inconsistent judgments further emphasizes the use of forward satisfaction procedures as a means of model improvement and domain understanding, as opposed to a decision-making tool. Although we have attempted to be neutral in our analysis, we are among the authors of several of the procedures under consideration. Our intention is not to declare one or more techniques as superior, but to understand the unique abilities of each procedure, helping goal modelers to select appropriate analysis approaches.

7 Conclusions

Analysis over goal models has been suggested as a means to aid in decision making using objectives captured in the model. In this work, we have provided a detailed comparison of forward satisfaction algorithms. We have applied seven sample procedures to three alternatives over three models using three tools. Results show that differing design choices and syntax interpretation among procedures can produce differing results over sample models. These results, if taken at face value, can lead users to select different feature or design alternatives captured in the model. The structure of the model to be analyzed plays a significant role in variance between procedure results and in the reliability of results in general. Specifically, the presence of many softgoals, contribution links, or dependencies may make results less reliable. Our results emphasize use of goal model analysis procedures should focus on their role as a tool to guide domain exploration, including model improvement, increased domain understanding, and improved communication, while their use as decision-making tool should be heuristic in nature. Future work could focus on studies that further test and qualify the benefits of goal model analysis in practice.

Acknowledgments Financial support has been provided by the Natural Sciences and Engineering Research Council of Canada andthe Ontario Graduate Scholarship Program.

Appendix 1

See Table 11.

Appendix 2

See Table 12.

Appendix 3

See Table 13.

Table 11 Results of the evaluation techniques applied for three alternatives over the Media Shop model in Fig. 1

	Alternative 1							Alternative 2							Alternative 3						
Element	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*
DB querying	*100	*FS	*100	*FS	*FS	*FS, N	*1.0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	0	N	0	N	N	N, N	0, 0
Catalogue consulting	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Pick available item	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Pre-order non-available item	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0
Classic communication handled	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0
Standard form order	0	N	0	N	N	N, N	0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Secure form order	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0
Monitoring system	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Produce statistics	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
System evolution	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Add item	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Check out	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Update catalogue	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Check authentication	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Check information flow	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Check access control	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Update GUI	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Shopping cart	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Select items	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Get identification details	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Internet managed	0	N	0	N	N	N, N	0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Perform privacy control	**90**	**PS**	**75**	**(FS)**	**(FS)**	**PS, N**	**0.5, 0**	**90**	**PS**	**75**	**(FS)**	**(FS)**	**PS, N**	**0.5, 0**	**90**	**PS**	**75**	**(FS)**	**(FS)**	**PS, N**	**0.5, 0**
Manage internet orders	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Adaption	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Manage internet searching	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Manage internet shop	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Privacy	**75**	**PS**	**75**	**FS**	**FS**	**PS, N**	**0.5, 0**	**40**	**N**	**50**	**(PS)**	**(C)**	**PS, PD**	**0.5, 0.5**	**40**	**N**	**50**	**(PS)**	**(C)**	**PS, N**	**0.5, 0**
Availability	25	PS	25	PS	**(FS)**	PS, N	0.5, 0	50	PS	25	PS	**(FS)**	PS, N	0.5, 0	50	PS	25	PS	**(FS)**	PS, N	0.5, 0
Integrity	**0**	**N**	**0**	**(C)**	**(C)**	**PS,PD**	**0.5, 0.5**	**0**	**N**	**0**	**(C)**	**(C)**	**PS, PD**	**0.5, 0.5**	50	PS	25	PS	**(C)**	PS, N	0.5, 0
Usability	**25**	**PS**	**25**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	**47**	**PS**	**24**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	90	PS	37	(PS)	**(C)**	PS, N	0.5, 0
Adaptability	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Easy to use	-25	PD	-25	PD	**(FD)**	N, PD	0, 0.5	-50	PD	-25	PD	**(FD)**	N, PD	0, 0.5	50	PS	25	PS	**(FS)**	PS, N	0.5, 0
Security	24	PS	25	(PS)	**(C)**	PS, N	0.25, 0	45	PS	19	(PS)	**(C)**	PS, N	0.25, 0	70	PS	25	(PS)	**(C)**	PS, N	0.25, 0

Bold values are significant

Table 12 Results of the evaluation techniques applied for three alternatives over the Wireless Services Model in Fig. 2

		Alternative 1							Alternative 2							Alternative 3						
Actor	Element	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*
Service Provider	Low Cost	37	PS	10	PS	**(FS)**	PS, N	0.25, 0	37	PS	10	PS	**(FS)**	PS, N	0.125, 0	**-12**	**N**	**-3**	**C**	**C**	**PS, N**	**0.125, 0**
	High Performance	**0**	**N**	**0**	**C**	**(C)**	**PS, N**	**0.25, 0**	**25**	**N**	**19**	**C**	**C**	**PS, N**	**0.25, 0**	**25**	**N**	**19**	**C**	**C**	**PS, N**	**0.25,0**
	Minimum Changes to Infrastructure	75	PS	38	(PS)	**(FS)**	PS, N	0.5, 0	**75**	**PS**	**38**	**(PS)**	**(FS)**	**PS, PD**	**0.25, 0.5**	**-25**	**N**	**-12**	**(C)**	**(C)**	**PS, PD**	**0.25, 0.5**
	Maximum Hardware Utilization	*50	*PS	*50	*PS	*PS	*PS, N	*0.5, 0	*50	*PS	*50	*PS	*PS	*PS, N	*0.5, 0	*50	*PS	*50	*PS	*PS	*PS, N	*0.5, 0
	High Throughput	**0**	**N**	**0**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	**25**	**N**	**19**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	**25**	**N**	**19**	**(C)**	**(C)**	**PS, N**	**0.5, 0**
	Minimum Message Exchange	100	FS	100	FS	FS	FS, N	1.0, 0	-50	PD	-25	PD	**(FD)**	N, PD	0, 0.5	-50	PD	-25	PD	**(FD)**	N, PD	0, 0.5
	Minimum Switch Load	-100	FD	-100	FD	FD	N, FD	0, 1.0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	PS, N	1.0, 0
System	Determine Data Location	100	FS	100	FS	FS	FS, N	1.0, 0	**0**	**N**	**0**	**N**	**N**	**FS, N**	**1.0, 0**	100	FS	100	FS	FS	FS, N	1.0, 0
	Data in Service Control Point	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0
	Data in New Service Node	0	N	0	N	N	N, N	0, 0	**-100**	**FD**	**-100**	**FD**	**FD**	**FS, N**	**1.0, 0**	100	FS	100	FS	FS	FS, N	1.0, 0
	Service in Central Switch	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	0	N	0	N	N	N, N	0, 0	0	N	0	N	N	N, N	0, 0
	Service in Service Control Point	0	N	0	N	N	N, N	0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
	Determine Service Location	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
	Install Service Node	0	N	0	N	N	N, N	0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Vendor	Service Nodes Ready for Sale	0	N	0	N	N	N, N	0, 0	*-100	*FD	*-100	*FD	*FD	*N, FD	*0, -1.0	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0
Depend-encies	Service Node	0	N	0	N	N	N, N	0, 0	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	*100	*FS	*100	*FS	*FS	*FS, N	*1.0, 0

Bold values are significant

Table 13 Results of the evaluation techniques applied for three alternatives over the Counseling Services model in Fig. 3

		Alternative 1							Alternative 2							Alternative 3						
Actor	Element	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*	*GRL-Quant*	*GRL-Qual*	*GRL-Hybrid*	*i**	*NFR*	*Tropos-Qual*	*Tropos-Quant*
Counsellors	Use Text messaging	*100	*FS	*100	FS	FS	FS, N	1.0, 0	*-100	*FD	*-100	FD	FD	N, FD	0, 1.0	*100	*FS	*100	FS	FS	FS, N	1.0, 0
Counsellors	Use Cyber Café/Portal/Chat Room	*-100	*FD	*-100	FD	FD	N, FD	0, 1.0	*100	*FS	*100	FS	FS	FS, N	1.0, 0	*100	*FS	*100	FS	FS	FS, N	1.0, 0
Counsellors	Provide Online Counseling Services	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0	100	FS	100	FS	FS	FS, N	1.0, 0
Counsellors	Help as Many Kids as Possible	50	PS	25	PS	**(FS)**	PS, N	0.5, 0	50	PS	25	PS	**(FS)**	PS, N	0.5, 0	50	PS	25	PS	**(FS)**	PS, N	0.5, 1
Counsellors	Listen for cues	**-50**	**PD**	**-75**	**(FD)**	**(C)**	**N, FD**	**0, 1.0**	**50**	**PS**	**75**	**(PD)**	**(C)**	**N, PD**	**0, 0.5**	-100	FD	-100	FD	(FD)	N, FD	0, 1.0
Counsellors	High quality counseling	-50	PD	-25	(PD)	**(C)**	N, PD	0, 0.25	**0**	**PD**	**13**	**(PD)**	**(C)**	**N, PD**	**0, 0.25**	-75	PD	-31	(PD)	**(FD)**	N, PD	0, 0.25
Counsellors	Avoid burnout	-25	PD	-6	PD	**(FD)**	N, PD	0, 0.25	-25	PD	-6	PD	**(FD)**	N, PD	0, 0.25	-25	PD	-6	PD	**(FD)**	N, PD	0, 0.25
Counsellors	Happiness [counselors]	**-12**	**PD**	**-1**	**(C)**	**(C)**	**PS, N**	**0.25, 0**	**13**	**N**	**8**	**(C)**	**(C)**	**PS, N**	**0.25, 0**	**-24**	**PD**	**-3**	**(C)**	**(C)**	**PS, N**	**0.25, 0**
Kids and youth	Use text messaging	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	-100	D	-100	FD	FD	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Kids and youth	Kids use cyber café/portal/chat room	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Kids and youth	Help be aquired	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Kids and youth	Comfortableness with service	**-25**	**N**	**-19**	**(PS)**	**(C)**	**PS, N**	**0.5, 0**	**-62**	**PD**	**-31**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	**0**	**N**	**0**	**(PS)**	**(C)**	**PS, N**	**0.5, 0**
Kids and youth	Anonymity [service]	50	PS	25	(PS)	**(FS)**	PS, N	0.5, 0	-25	PD	-25	(PD)	**(FD)**	N, PD	0, 0.5	**0**	**N**	**0**	**(C)**	**(C)**	**PS, PD**	**0.5, 0.5**
Kids and youth	Immediacy [service]	**-100**	**FD**	**-100**	**FD**	**(FD)**	**N, PD**	**0, 0.5**	**25**	**PS**	**25**	**FS**	**(FS)**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**(PS)**	**(C)**	**FS, PD**	**1.0, 0.5**
Kids and youth	Get effective help	**-37**	**N**	**-24**	**(C)**	**(C)**	**PS, N**	**0.25, 0**	**-31**	**PD**	**-8**	**(PD)**	**(C)**	**PS, N**	**0.5, 0**	**0**	**N**	**0**	**(PS)**	**(C)**	**PS, N**	**0.5, 0**
Organization	Use Cyber Café/Portal/Chat Room	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Organization	Use Text messaging	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**-100**	**D**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Organization	Provide Online Counseling Services	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Organization	Immediacy [service]	-50	PD	-25	(PD)	**(FD)**	N, PD	0, 0.5	50	PS	25	(PS)	**(FS)**	PS, N	0.5, 0	**0**	**N**	**0**	**(C)**	**(C)**	**PS, PD**	**0.5, 0.5**
Organization	Anonymity [service]	50	PS	25	(PS)	**(FS)**	PS, N	0.5, 0	-50	PD	-25	(PD)	**(FD)**	N, PD	0, 0.5	**0**	**N**	**0**	**(C)**	**(C)**	**PS, PD**	**0.5, 0.5**
Organization	Help as Many Kids as Possible	**-6**	**PD**	**-1**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	**-6**	**PD**	**0**	**(C)**	**(C)**	**PS, N**	**0.5, 0**	**-18**	**PD**	**-2**	**(C)**	**(C)**	**PS, N**	**0.5, 0**
Organization	High quality counseling	**-50**	**PD**	**-25**	**(FD)**	**(C)**	**N, FD**	**0, 1.0**	**0**	**N**	**0**	**(PD)**	**(C)**	**PS, PD**	**0.25, 0.5**	**-75**	**PD**	**-31**	**(FD)**	**(C)**	**PS, FD**	**0.25, 1.0**
Organization	Avoid scandal	25	PS	6	PS	**(FS)**	PS, N	0.25, 0	**-25**	**PD**	**-6**	**PD**	**(FD)**	**N, N**	**0, 0**	**0**	**N**	**0**	**(C)**	**(C)**	**PS, N**	**0.25, 0**
Organization	Increase funds	**-13**	**N**	**-4**	**(C)**	**(C)**	**PS, N**	**0.125, 0**	**-12**	**PD**	**-1**	**(FD)**	**(C)**	**PS, N**	**0.125, 0**	**-37**	**PD**	**-8**	**(PD)**	**(C)**	**PS, N**	**0.125, 0**
Organization	Help kids	**-28**	**PD**	**-6**	**(PD)**	**(C)**	**PS, N**	**0.25, 0**	**-3**	**PD**	**0**	**(PD)**	**(C)**	**PS, N**	**0.25, 0**	**-46**	**PD**	**-8**	**(PD)**	**(C)**	**PS, N**	**0.25, 0**
Dependencies	Help as Many Kids as Possible	**0**	**N**	**0**	**PS**	**FS**	**PS, N**	**0.5, 0**	**0**	**N**	**0**	**PS**	**FS**	**PS, N**	**0.5, 0**	**0**	**N**	**0**	**PS**	**FS**	**PS, N**	**0.5, 0**
Dependencies	High quality counseling	**-50**	**PD**	**-25**	**PD**	**C**	**N, N**	**0, 0**	**0**	**N**	**0**	**PD**	**(C)**	**N, N**	**0, 0**	**-75**	**PD**	**-31**	**PD**	**FD**	**N, N**	**0, 0**
Dependencies	Provide counselling via text message	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Dependencies	Provide counseling via cyber café/portal	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Dependencies	Text messaging service	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**
Dependencies	Cyber café/portal/chat room	**-100**	**FD**	**-100**	**FD**	**FD**	**N, N**	**0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**	**0**	**N**	**0**	**FS**	**FS**	**FS, N**	**1.0, 0**

Bold values are significant

References

1. Chung L, Nixon BA, Yu E, Mylopoulos J (2000) Non-functional requirements in software engineering. Kluwer, Norwell
2. Yu E Towards modelling and reasoning support for early-phase requirements engineering. In: Proceedings of 3rd IEEE international symposium on requirements engineering (RE'97), IEEE Press, pp 226–235
3. Dardenne A, van Lamsweerde A, Fickas S (1993) Goal-directed requirements acquisition. Sci Comput Program 20:3–50
4. Bresciani P, Perini A, Giorgini P, Giunchiglia F, Mylopoulos J (2004) Tropos: an agent-oriented software development methodology. Aut Agents MAS 8(3):203–236
5. Antón AI, Potts C (1998) The use of goals to surface requirements for evolving systems. In: Proceedings of the 20th international conference on software engineering (ICSE'98). IEEE Computer Society, Washington, DC, pp 157–166
6. Maiden NAM, Jones SV, Manning S, Greenwood J, Renou L (2004) Model-driven requirements engineering: synchronising models in an air traffic management case study. In: Proceedings of CAiSE'04, Springer, LNCS, vol 3084, pp 367–383
7. Horkoff J, Yu E (2009) Evaluating goal achievement in enterprise modeling—an interactive procedure and experiences. In: Proceedings of 2nd IFIP WG 8.1 working conference on the practice of enterprise modeling (PoEM'09), Springer, LNBIP, vol 39, pp 145–160
8. An Y, Gerrity P, Dalrymple PW, Horkoff J, Rogers M, Yu E Collaborative social modeling for designing a patient wellness tracking system in a nurse-managed health care center. In: Proceedings of the 4th international conference on design science research in information systems and technology (DESRIST '09), ACM Press, Article 2, 14 pages
9. Giorgini P, Mylopoulous J, Sebastiani R (2005) Goal-oriented requirements analysis and reasoning in the tropos methodology. Eng Appl AI 18(2):159–171
10. Amyot D, Ghanavati S, Horkoff J, Mussbacher G, Peyton L, Yu E (2010) Evaluating goal models within the goal-oriented requirement language. International Journal Intelligent Systems 25(8):841–877
11. Franch X (2006) On the quantitative analysis of agent-oriented models. In: Proceedings of CAiSE'06, Springer, LNCS, vol 4001, pp 495–509
12. Kaiya H, Horai H, Saeki M (2002) AGORA: attributed goal-oriented requirements analysis method. In: Proceedings of international requirements engineering conference (RE'02), IEEE Press, p 13
13. Bryl V, Giorgini P, Mylopoulos J (2007) Supporting requirements analysis in tropos: a planning-based approach. In: Proceedings of Pacific Rim international workshop on multi-agents (PRIMA'07), Springer, vol 5044, pp 243–254
14. Asnar Y, Bryl V, Giorgini P (2007) Using risk to evaluate design alternatives. Proc AOSE VII, Springer, LNCS 44405:140–155
15. Gans D, Schmitz D, Arzdorf T, Jarke M, Lakemeyer G (2004) SNet reloaded: roles, monitoring and agent evolution. In: Proceedings of international bi-conference workshop on agent-oriented information systems (II) (AOIS'04), Springer, LNCS, vol 3508, pp 68–84
16. Wang X, Lesperance Y (2001) Agent-oriented requirements engineering using ConGolog and i*. In: Proceedings of workshop on agent-oriented information systems (AOIS'01), pp 59—78. http://www.aois.org/
17. Fuxman A, Liu L, Pistore M, Roveri M, Mylopoulos J (2001) Specifying and analyzing early requirements: some experimental results. In: Proceedings of IEEE international requirements engineering conference (RE'03), IEEE Press, pp 105–114
18. Giorgini P, Massaci F, Mylopoulos J, Zannone N (2004) Requirements engineering meets trust management: model, methodology, and reasoning. In: Proceedings of international Itrust conference, Springer, LNCS, vol 2995, pp 176–190
19. Horkoff J, Yu E, Ghose A (2010) Interactive goal model analysis applied—systematic procedures versus Ad hoc analysis. In: Proceedings of 3rd IFIP working conference on the practice of enterprise modeling (PoEM'10), pp 130–144
20. Horkoff J, Yu E (2011) Analyzing goal models—different approaches and how to choose among them. Requirement engineering track—fourth edition, part of the 16th ACM symposium on applied computing, SAC'11
21. van Lamsweerde A (2004) Goal-oriented requirements engineering: a roundtrip from research to practice. In: Proceedings of IEEE international requirements engineering conference (RE'04), IEEE Press, pp 4–8
22. Proceedings of the i* Industrial Showcase (2011) JUCMNav (2011) http://jucmnav.softwareengineering.ca/. Accessed July 2011
23. Liu L, Yu E Designing information systems in social context: a goal and scenario modelling approach. In: Information systems, vol 29, number 2, April 2004, 2003, pp 187–203 (17)
24. Kaiya H, Horai H, Saeki M (2002) AGORA: attributed goal-oriented requirements analysis method. In: Proceedings international requirements engineering conference (RE'02), IEEE Press, p 13
25. van Lamsweerde A (2005) Requirements engineering in the year 00: a research perspective. In: Proceedings of the 22nd international conference on software engineering (ICSE '00), ACM, New York, pp 5–19
26. Mylopoulos J, Chung L, Yu E (1999) From object-oriented to goal-oriented requirements analysis. Commun ACM 42(1):31–37
27. Letier E, van Lamsweerde A (2004) Reasoning about partial goal satisfaction for requirements and design engineering. In: Proceedings of international symposium on the foundation of SE (FSE'04), ACM Press, pp 53–62
28. Tanabe D, Uno K, Akemine K, Yoshikawa T, Kaiya H, Saeki M (2008) Supporting requirements change management in goal oriented analysis. IEEE international conference on requirements engineering, pp 3–12
29. Maiden NAM, Lockerbie J, Randall D, Jones S, Bush D (2007) Using satisfaction arguments to enhance i* modelling of an air traffic management system. In: Proceedings of IEEE international conference on requirements engineering (RE'07), IEEE Press, pp 49–52
30. Giorgini P, Mylopoulos J, Sebastiani R (2004) Simple and minimum-cost satisfiability for goal models. In: Proceedings of CAiSE 2004, Springer, LNCS, vol 3084, pp 20–3–5
31. Horkoff J, Yu E (2010) Finding solutions in goal models: an interactive backward reasoning approach. In: Proceedings 29th international conference on conceptual modeling (ER'10), pp 59–75
32. Asnar Y, Giorgini P (2006) Modeling risk and identifying countermeasure in organizations. In: Proceedings of international workshop on critical information infrastructures security (CRITIS'06), Springer, LNCS, vol 4347, pp 55–56
33. Wang Y, McIlraith S, Yu Y, Mylopoulos J (2007) An automated approach to monitoring and diagnosing requirements. In: Proceedings of international conference on automated SE (ASE'07), ACM Press, pp 293–302
34. Horkoff J, Yu E (2009) A qualitative, interactive evaluation procedure for goal—and agent-oriented models. In: CAiSE forum, CEUR workshop proceedings
35. Giorgini P, Mylopoulos J, Nicchiarelli E, Sebastiani R (2002) Reasoning with goal models. In: 21st international conference on conceptual modeling (ER2002). Springer, Tampere

36. Bradel B (2009) Extending goal models with a probability model and using Bayesian networks. In: Proceedings of the international conference on software engineering research and practice, pp 543–549
37. JUCMNav (2011) http://jucmnav.softwareengineering.ca/. Accessed Mar 2011
38. OpenOME (2011) http://www.cs.toronto.edu/km/openome/. Accessed Mar 2011
39. G-R Tool (2011) http://www.troposproject.org/tools/grtool/. Accessed Mar 2011
40. Franch X, Maiden NAM (2003) Modelling component dependencies to inform their selection. In: Proceedings of COTS-based software systems, Springer, LNCS, vol 2580, pp 81–91
41. Franch X, Grau G, Quer C (2004) A framework for the definition of metrics for actor-dependency models. In: Proceedings of IEEE joint international conference on requirements engineering (RE'04), IEEE Press, pp 327–328
42. Bryl V, Giorgini P, Mylopoulos J (2006) Designing cooperative IS: exploring and evaluating alternatives. In: Proceedings of CoopIS 2006, Springer, LNCS, vol 4275, pp 533–550
43. Asnar Y, Bryl V, Giorgini P (2007) Using risk to evaluate design alternatives. In: Proceedings of agent-oriented software engineering (AOSE VII), Springer, LNCS, vol 44405, pp 140–155
44. Bryl V, Giorgini P, Mylopoulos J (2007) Supporting requirements analysis in tropos: a planning-based approach. In: Proceeding of the Pacific Rim international workshop on multi-agents (PRIMA'07), Springer, vol 5044, pp 243–254
45. Bryl V, Massacci F, Mylopoulos J, Zannone N (2006) Designing security requirements models through planning. In: Proceedings of CAiSE'06, Springer, LNCS, vol 4001, pp 33–47
46. Gans G, Jarke M, Lakemeyer G, Schmitz D (2003) Deliberation in a modeling and simulation environment for inter-organizational Networks. In: Proceedings of CAiSE'03 (LNCS), vol 2681. Springer, pp 242–257
47. Gans G, Jarke M, Lakemeyer G, Vits T (2002) SNet: a modeling and simulation environment for agent networks based on i* and ConGolog. In: Proceedings of CAiSE'02 (LNCS), vol 2348. Springer, pp 328–343
48. Wang X, Lesperance Y (2001) Agent-oriented requirements engineering using ConGolog and i*. In: Proceedings of the workshop on agent-oriented information systems, pp 59–78
49. Gans G, Jarke M, Kethers S, Lakemeyer G (2003) Continuous requirements management for organization networks: a (dis)trust-based approach. RE J 8:4–22
50. Gans D, Schmitz D, Arzdorf T, Jarke M, Lakemeyer G (2004) SNet reloaded: roles, monitoring and agent evolution. In: Proceedings international workshop on agent-oriented information systems (AOIS'04) (LNCS), vol 3508. Springer, pp 68–84
51. Fuxman A, Liu L, Pistore M, Roveri M, Mylopoulos J (2003) Specifying and analyzing early requirements: some experimental results. In: Proceedings international requirements engineering conference (RE'03), IEEE Press, pp 105–114
52. Fuxman A, Pistore A, Mylopoulos J, Traverso P (2001) Model checking early requirements specifications in tropos. In: Proceedings of IEEE international symposium on requirements engineering (RE'01), IEEE Press, pp 174–181