Comparing Representations of Contribution Labels in Goal Models

Wisal Tambosi

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE
STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN INFORMATION SYSTEMS AND
TECHNOLOGIES

YORK UNIVERSITY
TORONTO, ONTARIO

December 2019

Abstract

Goal models have been proposed to be an effective method to support decision making in early requirements engineering. Key to using them is the concept of contribution links that represent how the satisfaction of one goal affects that of another. Multiple proposals have been offered for representing contribution; however, the degree to which users can intuitively understand the meaning behind contribution representations and utilize them appropriately has not been thoroughly studied. This work reports the results of an experimental study that compares the intuitiveness of two contribution representation approaches by measuring the performance of untrained users and exploring the role of individual differences (cognitive styles and arithmetic attitude and ability) in establishing the right intuition. Results show significant differences between the two representations as well as effects of various levels of individual factors. The results inspire further research on contribution links and support the operationalizability of intuitiveness as a criterion for evaluating conceptual modelling language designs.

Acknowledgement

I would like to express my gratitude to my thesis advisor, Professor Sotirios Liaskos of the School of Information Technology at York University, who was my guiding light in my master's years from day one. Professor Liaskos helped me explore the various aspects of this field into finding the right path for my research. He was there to assist and advise me through every academic obstacle and challenge I have been through, setting me on the right track in my research methods, and making sure everything I do is thorough and precise. With his sincere mentorship, I managed to complete this thesis. I would also like to thank the committee members Dr. Zijiang Yang and Dr. Melanie Baljko for providing valuable feedback and interesting viewpoints on my research during the thesis defence.

At last but not least, I would like to dedicate this work to my dear parents, Yousif and Awatif, who supported me and never stopped believing in me, my beloved sister Ataa who was with me every step of the way, and all my family and friends, near and far, without whom I could have never made it this far or born the challenges that came along the road. Thank you for your continuous love, care and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the field of information systems analysis, conceptual models are graphical representations designed to facilitate understanding of the different aspects of the stakeholder needs and the end-system design. Such aspects include data flows, functional decomposition, entity structures and process flows. Goals models [7, 25, 74] are the type of conceptual model used to capture and assess the user's goals and intentions for a system design. For more than two decades, these models have been studied as an instrument for capturing and communicating such intentional structures for a variety of purposes within the field. One of the strengths of goal models is their ability to represent alternative ways by which stakeholder goals can be materialized into design solutions [41, 42, 53]. Using goal models, business/systems analysts can reason about and communicate the advantages and disadvantages of alternative solutions with respect to their impact to higher level business objectives.

1.1   Problem Statement

Goal models consist of a number of elements representing different kinds of goals (such as, soft goals, hard goals and tasks), as well as several relationships between them including decomposition, dependency and contribution. The latter kind of relationship, the contribution is of particular interest in our work. In goal models, contribution links indicate both the kind influence of one goal on another (i.e., positive or negative) and the magnitude of this influence (i.e., low or high). Several proposals for modeling contribution links have been presented in the literature, which can be distinguished between qualitative (symbolic) and quantitative (numeric). While each method has its own merits from an analytical standpoint, little is known with regards to how intuitive each method is for the users of the models.

The qualitative approach is the traditional approach for representing contributions. Symbolic labels (e.g."+', "–") or words ("help","break") express, in high-level terms, the quality (positive or negative) and the size of contribution. The numeric approach, on the other hand, use of numbers to indicate the contribution value, whereby, e.g., sign and absolute value are used to represent quality and size of contribution. Variations of both approaches have been proposed both with regards to representation and the underlying semantics.

Theoretical analysis and demonstrations are usually employed to support the soundness and usefulness of each approach. However, an additional indication of the quality of the chosen representation and semantics could be the extent to which untrained users of the model can *intuitively* understand the meaning of the representation and, moreover, use it to make inferences about the model in a way that complies with the semantics intended by the modelling language designers. Thus, the concept of intuitiveness can be regarded as an element when investigating the quality factors of conceptual

models. How the elements of each contribution link are visualized, and how correctly the meaning behind them is inferred by untrained users via their intuition, reflect the level of intuitiveness of that approach. Hence, when faced with a design choice with regards to the representation and meaning of a language construct, designers would opt for the alternative that aligns with what the intended users would find intuitive, at least all else being equal.

In this thesis, we explore the intuitiveness of two approaches for representing contribution links in goal models through an experimental design. The experiment core focus is two-fold. We first compare the two methods of contribution representations by presenting decision problems for users to solve. Second, we explore the role of individual differences in establishing the right intuition using either of the representation approaches.

For the first aspect of the experiment, a series of decision problems, modelled in either of the two ways (qualitative or quantitative) are presented to two groups of untrained users. The qualitative approach uses symbols (e.g ”+”, ”-”) are used to label contributions links indicating the influence of each goal over the other, while the quantitative approach uses numbers as labels. Users are asked to use the contributions to perform inferences and make decisions. We measure the extent to which their inferences comply with the semantics of each representation.

As a second aspect to our experiment, we look into individual differences of the selected users using different measures materialized in questionnaires, personality tests, and aptitude tests. Our experiment mainly explores how cognitive style, attitude and ability with mathematics and mental arithmetic as well as overall model use approach taken affect the degree of success in performing compliant inferences about goal models.

1.2 Research Objectives

Summarizing the above discussion, the study's primary objective is to to assess the accuracy of untrained users' perception of contribution link constructs in different goal modeling languages, and provide an educated comparison based on the results of the a controlled experiment. User's perception is also investigated through looking into the subjects' individual differences, measured using various measure. Particularly, the study has the following sub-objectives:

1. Review current goal modeling contribution link methods covered in the literature;

2. Review literature related to comprehensibility, intuitiveness and individual differences within the concept of goal modeling;

3. Investigate and compare the concept of intuitiveness within contribution modeling methods of goal models;

4. Measure and compare the efficiency of the two contribution links methods.

5. Comprehend the way in which people of different cognitive styles perceive goal modeling contribution links;

6. To address any correlation between users' arithmetic ability and attitude and their comprehension level of contribution links;

7. To compare the results of the above addressed questions and distinguish the method that presents a better intuitiveness for end users.

1.3 Research Questions

To achieve the objectives mentioned above, the following research questions (RQs) are addressed:

- **RQ1**: Which of the two methods for modeling contribution links (qualitative vs. quantitative) is better in terms of usability?

  RQ1.1: Which of the two methods for modeling contribution links (qualitative vs. quantitative) is more intuitive?

  RQ1.2: Which of the two methods for modeling contribution links (qualitative vs. quantitative) is more efficient?

- **RQ2**: How do individual factors affect the comprehension between the qualitative and quantitative methods of contribution links modeling?

  RQ2.1: Do cognitive styles affect the intuitiveness and efficiency of contribution link methods?

  RQ2.2: Does math anxiety affect the intuitiveness and efficiency of contribution link methods?

  RQ2.3: Does aptitude with mental arithmetic affect the intuitiveness and efficiency of contribution link methods?

  RQ2.4: Does the followed method and working approach affect the intuitiveness and efficiency of contribution link methods?

1.4 Thesis Contribution

We can summarize the main contributions of this thesis as follows:

- We offer a comparison between qualitative and quantitative contribution link approaches with regards to usability, comprehensibility and intuitiveness.

- We explore the notion of intuitiveness as one of the constructs for describing the effectiveness of conceptual models and show that its measurement is feasible and produces sensible data.

- We attempt a first exploration of individual differences in cognitive styles; mathematical ability; and mathematics anxiety, as influencing factors in the comprehension of goal models and the meaning behind the two approaches of contribution links.

- We effectively utilize Pinker's model of graph comprehension as a theoretical grounding of our work; understanding the cognitive processes involved in the process of conceptual model comprehension, and showing it can be used more widely in similar studies.

The result of this study will be valuable to the practitioners of information systems design and analysis as well as its end-users for the ultimate goal of keeping both parties in agreement on the meaning of goal modeling diagrams.

1.5   Thesis Organization

The thesis is organized as follows: Chapter 2 provides theoretical background on topics related to the field emphasizing the recent activities relevant to the subject. Chapter 3 reports the experimental design and the conducted procedures. Chapter 4 discusses the results of the experiments and their analysis, providing some further discussion on the implications of the study, future work and a summary. Finally, chapter 5 concludes the

work with final remarks and future work.

Chapter 2

Theoretical Background

This chapter provides a theoretical review of the main concepts of goal models and contribution link semantics. It also looks into the work that has studied the concept of intuitiveness, as well a the study of individual differences in conceptual modeling giving a brief background on the measures we take into account in our study. Finally, the chapter will look into research that have studied model comprehensibility.

2.1 Goal Models and Contribution Links

Goal models are conceptualizations of the intentional structure of a requirement analysis problem. They are believed to be useful in various stages of the information systems analysis and design life cycle, particularly requirements analysis, in that they allow representation of goals of various stakeholders and how they related to each other. Goal modelling frameworks proposed in the literature mostly share the same representational elements, but vary in the reasoning methods behind them. One of the most popular frameworks is the i* framework by Yu [74], which builds on ideas introduced earlier

with the Non-Functional Requirement (NFR) framework [15]. The name i* stands for "distributed intentionality" because the concept of i* is that it centres on the notion of intentional actors and intentional dependency, meaning that it shows how each actor depends on another actor for the achievement of its goals. GRL or Goal-oriented requirement language [2] is another framework that is based on the foundation of the i* methodology and uses a similar graphical notation in its models. GRL deals with three main concepts: Intentional elements, actors and links. The intentional elements consist of goals, resources, tasks and soft-goals. How GRL differs from the other frameworks [45] is that it provides constructs for expressing various types of concepts that appear during the requirements and high-level architectural design process. Following the introduction of GRL, User Requirement Notation or URN a was introduced, which combines the concepts of actors and intentions, from the GRL framework, with those of system behavior patterns from the Use Case Map (UCM) framework [7].

The goal models we focus on in this study are inspired by the i* modeling language, mentioned above, as it is one of the first and most widely known goal modeling notations. We particularly focus on Strategic Rationale (SR) diagrams within i* models, that is diagrams that explain why specific stakeholders have the goals they have by reference to higher level goals, and, reversely, how goals can be achieved through the achievement of lower level goals. The i* framework offers the ability to also develop Strategic Dependency diagrams (SDs) which show how actors depend on each other in order to achieve goals; such diagrams are outside our scope here and we instead focus on a version of SR diagrams. The SR diagrams we consider consist of elements of the following types:

1. Oval-shaped elements representing the functional goals, or hard goals – i.e. intentional objectives of clear satisfaction condition. Depicted in Figure 1, an ex-

ample of such goals is (*Choose Schedule*), because it represents a functional task to be performed which achievement can be checked.

2. Cloud-shaped elements representing the quality goals, or soft goals – i.e. objectives of imprecise satisfaction condition, such as performance quality, flexibility, or usability. In Figure 1, (*Minimal Conflicts*) is an example of this type of goals, because measuring the achievement of such goal cannot be accurately measured, and we can only estimate the degree to which (*Choose Schedule*) (*Manually*) or (*Automatically*) contribute to achieving it.

Apart from these goal elements, goal models also include relationships to express how knowledge of satisfaction of one goal, affects our belief about the satisfaction of other goals. These links are means-ends, which connect tasks to goals indicating a specific way to achieve the goal; task decomposition which indicate the sub-tasks, sub-goals, resources and soft-goals that need to be performed or satisfied in order for a task to succeed; and contribution relationships, which show how soft-goals contribute positively of negatively to achieving a quality, also show the AND/OR contributions. When it comes to task decomposition, analysts recursively decompose hard-goals into sub-goals to form an AND/OR decomposition which represent a fulfillment condition of a main goal by the sub-goals. An AND-decomposition indicates that the all sub-goals need to be satisfied in order for the destination goal to be satisfied as well. For example, (*Schedule Meeting*) goal in Figure 1 can only be satisfied when both goals (*Collect Time Tables*) AND (*Choose Schedule*) are satisfied. To the contrary, an OR-decomposition means that the satisfaction of at least one sub-goal results in the satisfaction of the destination goal. In Figure 1, (*Choose Schedule*) can be done either by '*Manually*' OR '*Automatically*'.

As opposed to mean-ends links and task decomposition links, contribution links
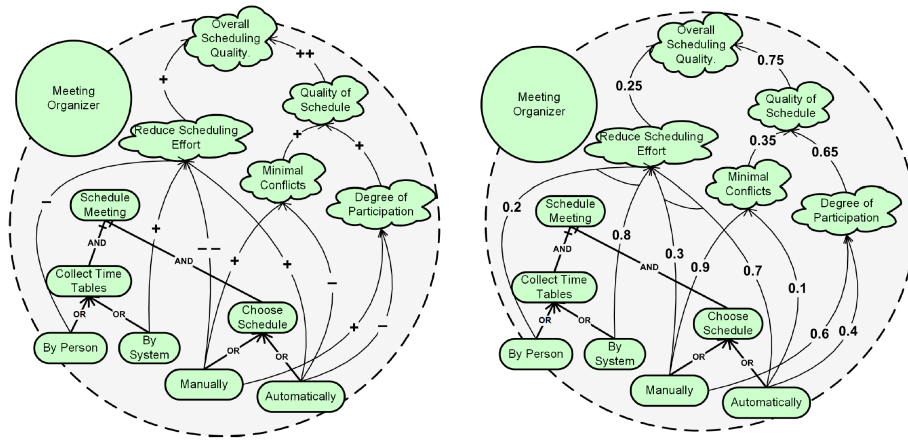
Figure 1: Goal models with symbolic (left) and numeric (right) contribution links.

are used when partial or uncertain satisfaction influence needs to be modeled. This is specifically useful for soft-goals whose satisfaction condition is, as we saw, not precisely defined. A contribution link from goal A to goal B means that goal A contributes to the satisfaction of goal B, i.e. the achievement goal A helps the achievement of goal B. The value assigned in the label of the connecting lines indicates the effect of whether it helps or hurts satisfying the attached goal of the upper level. Multiple studies have proposed different approaches on how these labels should be modeled and how to perceive their meaning [25] [26].

The qualitative or symbolic approach, used in modeling languages such as i* [74] and GRL [6], uses symbols such as "++", "+","−" or "−−" as labeling values to contribution links to indicate how the base goal affects the satisfaction of the destination goal. A goal model with qualitative contribution links is depicted on the left side of Figure 1. To provide an example, the task of *Collect Time Tables* can be achieved *By System*, which helps achieving *Reduce Scheduling Effort* goal, as inferred from the"+" symbol on the contribution link. An alternative is *By Person*, but hurts achieving *Reduce Scheduling Effort* goal, as inferred from the "−" symbol on the contribution link.

11

The precise meaning of qualitative labels, i.e. how they are supposed to be used to make inferences about satisfaction influence, has been discussed in various works. In [25] for example, specific rules that associate the satisfaction kind and level are set out and a logical reasoning technique allows rigorous (thus, amenable to automation) reasoning about propagation of satisfaction evidence. Elsewhere [6], a less expressive approach is adopted that however requires human intervention for deciding complex contribution combination scenarios. Both these efforts are explained below in Section 2.2.

The quantitative or numeric approach, on the other hand, use numbers, such as the values between [0,1], as labels to indicate the degree to which the base goal contribute in the satisfaction of the destination goal in comparison with its alternative [25]. Following the same example in the qualitative case, achieving the *Collect Time Tables* task by *By System* has the label value of 0.8, as opposed to *By Person* which has the value of 0.2, which indicates that the former participates better in the satisfaction of the *Reduce Scheduling Effort* goal than the latter. As with the qualitative case, various approaches for understanding the semantics of quantitative contribution links have been proposed in the literature. Giorgini et al [25] introduce a theory in which the modelers choose how the numbers are interpreted into satisfaction propagation following a general label propagation framework. In URN [6] as well as Liaskos et al. [40] a linear approach is presented which has simpler semantics but is less expressive and implies structural restrictions. These proposals will be discussed in more details in the following section.

2.2 Contribution Semantics

2.2.1 Overview

As we saw, the informal descriptions of contribution links above allow a model reader/user to perform very basic inferences by looking at the goal model. For example, a user can compare two contributions with respect to which one is larger or he/she can even choose between alternatives in the hard-goal decomposition with respect to a soft-goal of interest. Looking at Figure1, if to 'Reduce Scheduling Effort' is an important soft-goal, then we know that (*Choose Schedule*) 'Automatically' is preferable than doing so 'Manually' by simply looking at the contribution labels and without knowing precisely what they mean. However, more detailed semantics need to be given in order to perform more complex inferences such as deciding on the satisfaction status of a goal that receives multiple incoming contribution links, or, as we will see below, deciding the optimal alternative by considering all contribution links in the structure.

2.2.2 Qualitative Model Semantics

Giorigini et al. have developed the most expressive semantics for the qualitative (symbolic) links [25] [26]. According to their framework each goal in the diagram can be associated with two variables: one that measures satisfaction and one that measures denial. Each of these variables can take one of three values: Full evidence (denoted with prefix F), Partial Evidence (P) and No Evidence (N) of, respectively satisfaction (suffix S) or denial (D). For example, for a goal we may have partial evidence of satisfaction (denoted PS) and no evidence of denial (denoted here ND) and for another full evidence of satisfaction (FS) and partial evidence of denial PD; the inconsistency is perfectly fine and one of the strengths of the framework. A set of rules, seen in Table
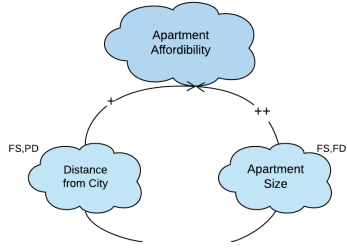
Figure 2: Qualitative Contribution
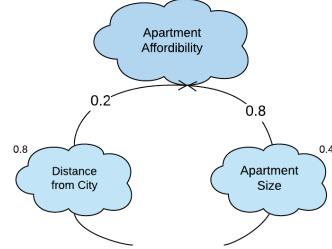
Figure 3: Quantitative Contribution

2.1, combine the satisfaction and denial values of the origin goal with the contribution symbolic label to decide the satisfaction and denial values of the destination. Looking at the example in Figure 2, 'Apartment Affordability' satisfaction value is calculated by looking at the criteria goals if we know that satisfaction and denial values of 'Distance from City' are FS , PD then based on the rules of Table 2.1 'Apartment Affordability' must be PS,PD – assuming no other influence. However, considering the other criteria 'Apartment Size' satisfaction value FS , FD, the maximum of both criteria value will be FS , FD Which is 'Apartment Affordability' satisfaction and denial value.

Table 2.1: Symbolic Contribution Semantics

Label	Effect	Label	Effect	Label	Effect	Label	Effect
++	FS → FS PS → PS PD → PD FD → FD	−−	FS → FD PS → PD PD → PS FD → FS	+	FS → PS PS → PS PD → PD FD → PD	−	FS → PD PS → PD PD → PS FD → PS

Other proposals for the semantics of contribution labels have also been proposed in the literature. Amyot et al. [6] use a single value system and as such use a more complex function that explicitly labels conflict. Which is similar to the previous proposal in a sense that the strength of the contribution effect is the minimum between the strength of the label and the satisfaction of the origin, noting that negative labels invert satisfaction into denial and denial into satisfaction. Aggregation however is different

Table 2.2: Numeric Contribution Semantics

Bayesian	$s(g) = \underset{g' \in O_g}{MAX}\{s(g') \times w(g',g)\}$
Min-max	$s(g) = \underset{g' \in O_g}{MAX}\{MIN(s(g'), w(g',g))\}$
Serial-parallel	$s(g) = \underset{g' \in O_g}{MAX}\{\frac{s(g') \times w(g',g)}{s(g') + w(g',g)}\}$
Linear	$s(g) = \sum_{g' \in O_g}\{s(g') \times w(g',g)\}$

in Amyot et al. where strong and weak effects are counted and compared separately to then combine in a hybrid addititive/maximization fashion marking co-presence of strong positive and negative effects with "conflict" labels. We note that, in the context of such conficts, Horkoff et al. [33] suitably proposes human intervention for their resolution, instead of relying on rules.

2.2.3 Quantitative Model Semantics

In the quantitative (numeric) framework the rules are replaced by algebraic formulae. Based on the work of Giorgini et al. [25] [26], three possible ways were proposed by which this formula can be structured, seen in the top three rows of Table 2.2; in practice, their framework is open to the adoption of many other ways. Given a set of goals $g' \in O_g$, each with satisfaction value $s(g') \in [0.0, 1.0]$ targeting goal g with contribution links weighted as $w(g',g)$, the satisfaction value of goal g is expected to be $s(g)$ as defined in each of the formulae. In all the proposed formulae (*"Bayesian", "Min-Max"* and *"Serial-Parallel"*) aggregation is implemented through maximization. Note that in this semantic framework, users are supposed to understand the numbers of the contribution links as absolute contribution values potentially elicited and understood in isolation from the other ones.

A different interpretation of numeric contributions, which is of particular interest

here, is the approach followed by URN [6]; a version of which was experimentally studied by Liaskos et al. [40]. According to that interpretation, a unique numeric satisfaction value is assigned to each goal with values in the real interval [0.0,1.0] showing no distinct satisfaction and denial values. Then, the number on the contribution link denotes the share of contribution of the satisfaction of the origin goal to the satisfaction of the destination goal. This implies also a different formula for satisfaction propagation, the last one on Table 2.2; the formula is labeled as *"Linear"* because it calculates the satisfaction of the destination goal through linearly combining the satisfaction value of each goal that influences it, using the numbers on the contribution links as weights for the linear combinations. To provide an example, a goal model depicted in Figure 3 show that 'Distance from City' has the satisfaction value of 0.8, while 'Apartment Size' has the satisfaction value of 0.4. When calculating the satisfaction value of 'Apartment Affordability' following the Linear Formula will be 0.48.

Quantitative frameworks use algebraic expressions instead of rules and exhaustive tables. Amyot et al. [6] multiply satisfaction values of goals A_i (a number in [-100,100]) with the label l_i (also a number in [-100,100]). The satisfaction of B is calculated by adding up the results – as in $sat(B) = \sum sat(A_i) \times l_i$. We can, thus, say that, given the resulting satisfaction value for B, the labels of the incoming contributions represent the *share* of each of the corresponding origin goal to the satisfaction of B. In Liaskos et al. [40] and Maiden et al. [47] the "share" semantics becomes more explicit: labels are the result of Analytic Hierarchy Process (AHP) process and must necessarily add up to 1.0. As opposed to the URN approach, however, each goal can receive multiple groups of incoming contribution links, each group independently concerned with a specific local decision. Thus, this AHP-based framework is not concerned with calculating a global satisfaction value that results from a total evaluation of a goal model, but rather

16

sets of satisfaction values corresponding to options in decision problems expressed as OR-decompositions in the model. Another important difference of that framework is that it does not define denial of goal, which greatly simplifies effect and aggregation rules.

Despite the availability of both approaches, it is observed that the qualitative representation is more widely used, for reasons that can be hypothesized to relate to their perceived appropriateness with respect to the data that exist at hand about the contribution relationships which are often limited, e.g. in early requirements analysis contexts. For example, considering a goal "Distance from City" and the goal "Apartment Size", we can convincingly argue that the former contributes "positively" or "very positively" to the latter and represent it through a "+" or a "++" respectively. It is arguably less convincing to say that the contribution is 0.2 or 0.8, as it raises the question of where the numbers came from.

Nevertheless, Liaskos et al. [40] showed that adopting the linear interpretation of contribution links comes with simplifications and restrictions. This interpretation offers the benefit of systematic elicitation of the numbers through AHP and the pair-wise comparison method it introduces. Following this approach, numbers such as 0.2 and 0.8 are actually not arbitrary but the results of a streamlined and widely-applied elicitation process [47]. This way, the arbitrariness argument against quantitative goal models is addressed. What remains to be understood is how qualitative labels and numbers compare as visual representations. In other words how the choice of representation affects the degree by which users/readers of the models comprehend and utilize its content.

2.3 Theory of Graph Comprehension

Goal models are conceptual models, i.e. conceptualizations of a domain of interest. While a conceptual model can be represented in different ways, the dominant approach is through the use of the diagrammatic representations such as those we discussed earlier (clouds, ovals, links, labels of various kinds). However, successful understanding of a conceptual model necessarily depends on the choice of the appropriate diagrammatic (or other) representation, which, in turn, is informed by our understanding of how humans perceive such diagrams. In this section we focus on this question of graph comprehension.

There are different methods to represent and deliver information. Verbal or sentential approach is the basic method when communicating an information in confrontation between two parties; however, visual graphs are also used and have been proven to be an effective method to communicate information and facilitate its delivery because they utilize cognitive and perceptual mind mechanisms effectively [38].

Several theoretical interpretations on the mental processes involved in the phenomena of graph comprehension have been studied. The work we present in this thesis is strongly based on Pinker's Theory of Graph Comprehension [55]. This theory by Steven Pinker explains the cognitive operations and processes involved in reading an information graph, predicting through them, the traits that makes an individual better or worse at reading a graph. In understating these processes, the theory in return predicts the graph properties that take part in its ability to convert a given type of information to the reader.

In its essence, graphs translate information into visual objects, and according to Pinker's theory, the dimensions of each objects correspond to a mathematical scale in terms of its location and relationship with other objects. Assigned values to each object

correspond to a value on that mathematical scale. To understand these visual elements and their correlation, Pinker explains a number of mental processes the reader goes through. In what follows, the concepts used in the theory to reflect the outcomes of these mental processes.

- *The Visual Array*: The term reflects the result of the first mental reception of the graphic display, which is the raw information the reader should reason with using an encoding mechanism.

- *The Visual Description*: Through the reasoning process mentioned above, the Visual Description is created, which encodes the marks depicted on the graph based on their physical dimension using words and symbols.

- *The Graph Schema*: Through the recognition of the type of graph the reader is faced with, the reader instantiates what is called a Graph Schema; which shows how the physical dimension is mapped onto the appropriate mathematical scale.

- *The Conceptual Question*: This term represents the message the graph is meant to deliver which the reader is trying to assemble. In case the extracted message was not the intended conceptual question, inferential processes of mathematical reasoning are performed to infer the correct information.

Based on the theory concepts mentioned above, all sorts of information graphs should trigger the same cognitive processes in order to be comprehended. Using the above-mentioned mental methods to reason with the encoded visual description and graph schema, the reader then tries to achieve comprehension by predicting the conceptual question of the graph. Pinker indicates that individual factors have an influence on the accuracy level of the information retrieved by the reader.

19

To provide an example using goal models, we can assume that the first information received when exposed to a goal model like such as those in Figure 1, i.e the notice of the mere existence of shapes, lines, and texts represent the *Visual Array* of the graph. Once realizing the elements of the graph, the reader begins to create a Visual Description in which the mind makes connection of how each element is connected and related to others in terms of pattern location, distance, and included data without yet noticing the rules, and reasoning behind these elements. In the next step, the reader of the graph will search in their long term memory for a suitable Graph Schema, a template, that is, for interpreting the incoming visual information. The Graph Schema holds the information necessary to assign meaning of the visual elements and perform inferences therewith. To be effective as such the Schema is customized to the specific instantiating at hand. For example, when a reader sees a flow-chart, the Graph Schema of flowcharts is evoked, in which the boxes and lines allow for specific reasoning. Subsequently the schema is customized for the flow-chart at hand. Which Graph Schema is evoked is highly driven by a pattern matching process between the visual information and similar representations observed in the past.

Other Theories

Although our theoretical basis is Pinker's, various other theories of graph comprehension, largely compatible with it have been proposed in the literature that are worth mentioning.

In a different work by Kosslyen [37], A take on the matter of visual information processing is summarized. In his paper on understanding graphs and charts, the visual array explained by Pinker represents the perceptual image in Kosslyn's work, which is an output of what the reader visually make out out of a given graph. This image

remains so as long as it remains within display. The material of that image is then stored in the short-term memory as basic information accompanied by conscious experience when recognizing certain words or symbols in the graph. This shot-term information is connected with information received from the long-term memory which would aid conferring meaning beyond the given graph.

Kosslyn also studies the efficiency of graphs using an analytic scheme that evaluate graphs on four levels. The first level is the *Basic Level* in which the graphic constitutes (background, framework, specifier, and labels) are defined. On the *Syntax Level*, each of these constitutes is evaluated as visual elements based on their location, dimension, and grouping. The *Semantic Level* then analyses the meaning behind each element and what they represent and depict in the graph. Finally, the *pragmatic level*, which analyzes how each element is interpreted by readers beyond what they visually represent. This analytic scheme is used to evaluate existing graphs in terms of their efficiency, but also suggested to be used and applied in the design process of graphs.

Another theoretical approach pertinent to graph comprehension is due to D.Moody [51] who has expressed the lack of focus on visual syntax in favour of semantics when developing a visual notation, and the lack of guidelines in the literature with regards to visual representations. To that end, he proposes the Physics of Notation framework in which he provides nine principles purposed to evaluate and compare existing notations, or use as a guideline for designing a cognitively effective visual notation optimized for human communication and problem solving. He further uses these principles to identity flaws in some existing software engineering notations, suggesting points to improve them. In what follows a summary of of these principles:

- Semiotic Clarity; which is concerned with the correspondence between the semantic constructs with the graphical symbols of the designed notation.

- Perceptual discriminability; which relates to the distinguishability of different graphic symbols.

- Semantic transparency in which graphic representations should be inferable. Semantic Transparency is the notion that is the closest to our notion of intuitiveness discussed below.

- Complexity Management which refers to the ability of visual notations to deliver information without overloading the human mind.

- Cognitive Integration when multiple graphical diagrams are used to represent a system.

- Visual Expressiveness; to evaluate the number of visual variables (shape, texture, brightness, size and color) used, and their utilization.

- Dual Coding; which is concerned with whether or not the use of text alongside graphics help with their interpretation.

- Graphic Economy; which related to the number of symbols used in the notation and whether is cognitively manageable.

- Cognitive Fit; which follows a theory of the same name in which different graphic representations of the same information are used for different tasks and audience.

Within the domain of our research, the principle of semantic transparency pointed out the concept of intuitiveness of the visual representation. According to that principle, there should be an association between graphic representations such as symbols and relationships and the concept they represent or indicate in order to ease the cognitive load on the human mind.

2.4 Intuitiveness

Given our above understanding of graph comprehension as a process of evoking and instantiating a schema, i.e. a framework for understanding the various elements in the visualization and how they are supposed to work, it follows that the process may be successful or unsuccessful under different circumstances. In our work we are particularly concerned with whether the right schema and rules of inference are evoked by the visualization, when users are not provided any training for doing so. To develop such a construct, we borrow the term intuitiveness used in daily life to refer to properties of functions (e.g. of devices, human-machine interfaces) that makes them usable without training.

Intuitiveness is understood as having the ability to know or understand things without any proof or evidence. We use the (working) theoretical construct "intuitiveness" of a model construct to describe the ability of untrained users of a conceptual model to readily understand what the construct means and how it should be used to make inferences in the model. The concept is analogous to the idea of an intuitive human-machine interface: the more intuitive an interface is, the more readily first-time users can use it without the need to resort to help, a manual etc. The term is akin to that of *learnability* which in its general term refers to the ease in which something can learned. Within system and software design, it is a quality of an interface that allows users to learn how to use it easily and quickly [66], the ease of this learning process explained in the match between the effort that users dedicate to learn a system one one hand, and whether they become effective in that on the other. One can think of intuitiveness as a facilitator of learnability in interactive visualizations, both being finer quality factors to usability. Design principles such as consistency and compliance to standards [54] are understood here to facilitate intuitiveness: users will likely find intuitive a user in-

terface that uses conventions with which the user is already familiar. We note that in the study we describe in this thesis, the training offered to the participants refers to the context in which the object of intuition, i.e. the operational meaning of contribution labels, is being investigated rather than the object itself. In other words, we neither train participants to specific contribution semantics, assessing afterwards the outcome of the training, nor do we observe the level of improvement as they perform tasks, as feedback is not provided and such "improvement" is not expected. This makes us hesitant to characterize our focal construct as learnability, despite it being a more established construct in the field of Human Computer Interaction.

With this user-machine interface analogy in mind, we can reasonably claim that conceptual models are also artifacts to be efficiently used by people, where "use" here is "understanding and communication" [53]. Further, as design artifacts themselves, modeling languages are results of design decisions at two levels: at the level of the concepts they consider (e.g., hard-goals and soft-goals) and at the level of the visualization of those concepts (e.g., ovals and clouds). It might be the case that there are better and worse decisions for each of those levels.

Intuitiveness, as applied in this work, measures the entire package of a concept and its visualization: the visualization evokes a meaning, which, in return, is used to make inferences. When a user is exposed to a visualization and ends up performing an inference that is not intended by the designers, a sub-optimal decision may be claimed at any of the levels: either the users did not map the visualization to the right concept (e.g. confused a "goal" for an "event", both otherwise being clearly understood concepts), or they did so correctly but did not understand the concept as the language designers intended them to (e.g., they correctly mapped a symbol to an "upper-goal" but did not know what to do with the latter). While training may arguably establish

24

correct bridging between visualization and inference in the long term, intuitiveness is exhibited when limited such training is necessary.

In the context of contribution links in goal models, the inference this work is interested in is how users assign satisfaction to goals given satisfaction of other goals based on their own intuition and interpretation of what contribution labels probably mean. Reversely, their observed inferences are revealing of their perceived meaning of the links, and, as such, the former can be used as empirical operationalizations of the latter.

2.5 Individual Differences in Model Comprehensibility

When performing tasks such as reasoning with diagrams, humans employ, as we saw, a combination of cognitive processes. It is natural to assume that not all individuals will employ these processes in the same way, combinations, efficiency and accuracy, and, moreover, these differences will likely affect the outcome of the task. The subject of individual differences in modern psychology refers to the psychological differences and similarities between people, which may, in our case, explain different levels and qualities of performance when reasoning with diagrams. The methods that are concerned with identifying, measuring and classifying these differences are developed within the context of psychometrics, i.e. psychological assessment [72]. Within the field of Information Systems, individual differences have been investigated as influencing factors in different areas; we offer some examples here. Harrison et al. [29] studied the influence of demographics (gender, age, education and experience), personality (computer attitude, computer anxiety and math anxiety) and cognitive styles on skill in End-User computer. In the same manner, Sein et al. [63] studied the effect of visual ability and

learning mode on the mental model formation process in novice users of a computer system. Humpherys et al. [35] presented a model to related personality dimensions, motivation and anxiety to information processing.

Within the context of conceptual modeling, a number of individual characteristics such as age, gender, education, domain expertise, systems experience, skills, and cognitive style have been identified as important variables in the design and implementation of information systems. In his study, Dhillon [19] has considered this approach to study how individual differences in conceptual modelers can influence the task performance of conceptual modeling. The results have shown a probable relation between the cognitive styles factor of modelers and the conceptual models quality, while self-efficacy seem to be a direct determinant of the model quality.

Since one of the concerns of this study is to look into model comprehensibility, explained later in section 2.6, Individual factors have also been studied within that context. For example, Reijers and Mendling [58] investigated the personal factors that could affect the comprehensibility of process models using self-assessment surveys on a group of students to assess their theatrical knowledge and expertise in process models, which comes with agreement to the study of Shanks et al. [65]. Mendling's work [49] has shown that there is a connection between personal factors, i.e. the background knowledge of conceptual models, and the ability to understand process models. In comparing between two modeling languages, De Lucia [46] found that the subjects experience and ability were of influence when the more experienced subjects performed better than the novice ones.

For our research we are concerned with whether cognitive individual differences affect the success and ease by which users can use goal models to make decisions in either the qualitative or quantitative models. We will particularly focus on cognitive

style skills and mathematical skills and aptitude.

2.5.1 Cognitive Styles

The cognitive method or style in its concept refers to the way in which different individuals think, perceive and remember information. Cognitive Styles influence the way we receive, organize, and interpret the data of our surroundings, and how we then integrate these interpretations into mental models that guide our behaviour and influence the cognitive tasks we perform [29]. The value of cognitive style has been increasing in recent years. The concept and its measures has been utilized for selection purposes such as careers guidance, task design, team composition, conflict management and training, even on a personal level [30]. Individual differences in cognitive styles have been regarded as important in relation to influencing perception, learning, decision making, communicating and information processing [50]. Xue et al [73] studied their impact on the decision making choices of firm managers, and has recognized them to be a major factor.

In most literature, cognitive style is understood as a bipolar construct revolving around the dimension of analytical style versus intuitive style. Intuitive style is characterized by knowing without knowing the reasoning beyond the knowledge of the made decision. Analytical style, on the other hand, is the knowledge resulting from breaking down the problem and collect information in a systematic explainable method [4]. Most work tend to categorize individual cognitive method of dealing with information as leaning towards one of the cognitive styles as a personality trait [30]. However, other work label cognitive styles differently. When Cools and Broeck [16] proposed their Cognitive Styles Indicator (CoSI), they labeled three cognitive styles: knowing, planning, and creating to identify how different individuals deal with information and

their problem solving methods. Martin's Cognitive Style Inventory [48] follows the concept of cognitive style model which uses consisted of two continua: (1) high systematic to low systematic and (2) high intuitive to low intuitive, which result in five different styles depends on the score location in the continua: systematic style, intuitive style, integrated style, undifferentiated style, and split style. This study, however, uses Allinson and Hayes' 38-item self-report Cognitive Style Index (CSI) [4], a measure specifically designed for survey purposes. The CSI measure individuals based on their rationality versus intuitiveness tendency in handling decision making, which is a main concern of this study.

When comparing between the cognitive methods, especially those dividing them into analytical versus intuitive styles, Hammond [28] states that research has been in favor of the intuitive approach being more effective when dealing with a decision making situation. In his work, however, Hammond argues that these comparison has been rather restricted mainly due to the use of analytical methods to measure intuition, and he states that both intuition and analysis modes can be triggered within the same person based on the assigned task property. He offers a method to indirectly compare between the two modes using the concepts of *cognitive continuum* and *task continuum*, which purpose is to determine whether the assigned task is intuition-inducing, or analysis-inducing. The theory works when the properties of the task at hand is identified through the attributes it represents, i.e how it is presented, displayed, whether it can measured, or should follow a certain method to be solved. This, in return, induce one of the cognitive modes. He identified the cognitive properties of both intuition and analysis, presented in Table 2.3, when handling a given task. A simple example is how presented mathematical problems usually induce analysis, while graphical displayed problems induce intuition. However, the task properties can be displayed on the depth

Table 2.3: Properties of Intuition and Analysis

	Intuition	Analysis
Cognitive control	Low	High
Rate of data processing	Rapid	Slow
Conscious awareness	Low	High
Organizing Principle	Weighted average	Task specific
Errors	Normally distributed	Few, but large
Confidence	High confidence in answer; low confidence in method	Low confidence in answer; high confidence in method

level, which refers to the covert relationship among the variables; or the surface level.

or the surface level, which refers to the over display of the task variables. In case of the lack of congruence between the surface and depth level, the apparent or surface properties of a task may appear to induce a certain cognitive mode, while its depth properties actually induce another. For example, the task of appreciating the aesthetic features of mathematical formulas. His work concluded that when specifying task properties on both depth and surface level help predicting the cognitive process used to solve that task.

2.5.2 Mathematical Aptitude

Mathematical aptitude has always been a case of interest mainly in the educational settings. In the literature, it was investigated as a dependent factor with multiple factors of influence such as gender, cognitive abilities, and anxiety [20]. The use of mathematical ability to make decisions and solve problems falls under the concept of quantitative reasoning. There is a strong belief that mathematical skills are crucial in almost in area of practice due to the large number of studies showing strong correlation between mathematical performance and workplace performance. This belief lead employment to seek a certain level of mathematical aptitude in individuals for job and task assigning purposes. A large number of aptitude tests developed by psychometric experts exist in

various organizational resources and online such as SHL Tests, Talent Q and Cubiks provided through JobTestPrep [3]. However, the availability of these tests is exclusive to personal and employment purposes.

Other than the area of business and employment, math assessment tests are also used in education to identify math difficulties in students. An example of such is TOMA-3 (Test of Mathematical Abilities - Edition Three) [11], which contains four sets of tests assessing mathematical knowledge, computation, math in everyday life, and word problems.

In this work, we seek to study mathematical aptitude as a psychometric measure in our experiment. In particular, we developed a mathematical aptitude test to focus on arithmetic branch of mathematics that is concerned with the the ability to perform the traditional operations of addition, subtraction, multiplication and division of decimal numbers which other tests does not normally offer. The purpose of this measurement is to find correlation between mathematical aptitude and the ability to infer optimal choices when exposed to goal models of quantitative methods.

2.5.3 Mathematics Anxiety

We intend to study mathematical anxiety as an individual difference factor of impact on the perception of contribution links in goal models. Given the nature of analytical structure of conceptual models and goal modelings in general, and the quantitative approach of contribution links in particular, we hypothesize that mathematical anxiety would have an impact on the level of perception in the experiment subjects.

Math anxiety is a state of mind affects the motivation to learn mathematics or solve math-related problems in both academic and ordinary life [21]. Ashcraft [8] defined math anxiety as "a feeling of tension, apprehension, or fear that interferes with math

performance". In his work, Ashcraft discussed the personal, educational and cognitive consequence of math anxiety, and found that within the context of decision-making, math anxious individuals tended to make more errors as the problem became more complex when compared to low-anxiety participants. Harrison et al. [29] studied the influence of math anxiety on end-user computing skills, and found that there is a negative relationship between the two.

Multiple measures were proposed in the literature to assess math anxiety. Richardson and Suinn [60] devised Math Anxiety Rating Scale (MARS), which consists of 98 items reflecting a variety of life situations that adolescents might experience that would involve dealing with numbers. An shortened version of the same measure was later developed by the same authors that contains only 30 items, which was validated to be comparable to the original [69]. A revised 24-item test (MARS-A) by Plake and Parker [68] has also been proposed. In our study, we adopt the 9-item Abbreviated Math Anxiety Scale (AMAS) by Hopko et al. [32].

2.6 Related Empirical Work on Model Comprehensibility

We close this chapter by looking at related empirical work that has been done in the literature on constructing and assessing comprehensibility. The concept of comprehensibility within the field of conceptual modeling has been a focus of several studies in the past few years. As a criterion, it is important to conceptual models in order for them to support the communication and provide understanding about the structure and functionality of an information system during the engineering phases and beyond.

Much of the research in the field has been dedicated towards understanding the

comprehensibility of (various aspects of) Unified Modeling Language (UML) and Entity-Relationship (ER) diagrams – e.g., Shovel and Frumermann [67] conducted an experiment to compare between the Extended Entity-Relationship (EER) model, and the Object-Oriented (OO) Model to investigate which of the models is easier to comprehend. Although no major differences has been spotted in the results, it however reveals some of weaknesses in the OO models, and some usability factors of EER. In the same manner, De Lucia et al. [46] compared the two modelling languages of ER (Entity-Relationship) and UML (Unified Modeling Language) through a controlled experiment to investigate which of the two languages is more comprehensible. Factors investigated as variables in the study were models designed in both UML and ER, comprehension level, subjects' experience, and subjects' ability. The Results have found that UML is easier to understand than ER diagrams. Performances of studied tasks using UML show higher results than those using ER. The subjects experience and ability were also of influence using UML given graduate students subjects performed better then undergraduate students, but this does not apply to performances on ER diagrams for there was no significant difference between the two subject groups there.

When dealing with notations in the same modeling language, Cruz-Lemus et al. [18] reviewed a number of studies on UML statecharts, developing an experiment to study the effectiveness and comprehensibility of these statecharts. Their results showed minor conflicts, but overall stated the benefits of statecharts usage being situational. In a different study, Purchase et al. [56] demonstrated the usage of an experimental framework to compare between ER diagrammatic notations, and involving users to investigates their preference in terms of comprehensibility. The results showed that such methodology is usable to test the comprehensibility of diagram notations.

A number of studies have focused on process model comprehensibility where a

number of design factors have been considered. When the structure of the conceptual model is of concern, the modeling Language used for designing the model show no major influence on comprehensibility as studied in [57] [46]; however, in a study to compare the usability of Business Process Model and Notation (BPMN) and Unified Modeling Language 2.0 Activity Diagram (UML AD), Birkmeier et al [9] conducted an experiment in which they assign trained users to create models using the mentioned modeling notations. The results have found that UML AD has exceeded BPMN on the criteria of better data handling. In addition, their work have revealed that BMPN has technical limitations concerning flexibility. Figl et al [24] introduced a number of evaluation principles when model design is of interest. Their study was concerned with the criteria symbols represented in process models, and how to evaluate their effectiveness when it comes to model comprehension.

In a different study, Figl and Laue [23] studied the impact of BPM elements design factors, i.e their relations, interactivity and separation, on the comprehensibility of these elements. The result showed interactivity having the bigger impact, and concluded that reducing the cognitive load in general when creating process models improves their understandability; which in most cases depends on the nature of the domain. This comes to an agreement with the study of Shallas et al [62], which shows that complexity of the conceptual models have a negative correlation with how well a model is perceived. The higher the complexity, the more challenging it was for users to comprehend.

Modeling design structure-wise, Mendling and Strembeck [49] has found that structural and textual factors related to the process model purposes, characteristics and layout strategies influence the understandability of process models. Another structural factor is the size of the model and the load of information it contains, which was also

investigated by Reijers et al. [58]. Their results emphasized the need for structural guidelines for understandable process model, and also the need for the training to create such models.

Although the comprehensibility of models or *understandability* is a popular construct of study, it has been argued that there is little agreement on how this is to be measured. Indeed, in their survey, Houy et al. [34] find variability in how understandability is operationalized in the literature. The concept of intuitiveness, as a specialization of understandability, is less frequently being focused on explicitly as in the work by Jošt et al. [36], for example, where the *intuitive understandability* of various methods for modeling processes are empirically compared. The modeling languages involved in the study were UML AD, BPMN, and Event-Driven Process Chain (EPC). Results has shown that BPMN seems to surpass when processes are of minor complexity, while EPC proved to be more beneficial for higher process complexity. However, the study has also shown neither these languages significantly outperformed UML AD, which makes it the most versatile choice for intuitive comprehensibility.

A number of metrics have been proposed by to ensure the comprehensibility and overall quality of conceptual models. Vanderfeesten [71] introduced Cross-Connectivity as a metric to measure the relationship between process model elements (nodes, connectors, and arcs) to test its comprehensibility and accuracy. The base of the introduced metric is considering the cognitive efforts of the model receiver in order to understand it. Serrano at al. [64] conducted an experiment to validate a number of proposed metrics that aids in assuring the quality of models used to interpret data retrieved from data warehouses. Those metrics assess some quality aspects of the conceptual models including different scopes and levels of the model's elements. His work has shown correlation between some of the metrics and the comprehensibility aspect of the con-

ceptual models.

Goal Model Comprehensibility

Work that relates to understanding the comprehensibility of goal models specifically is more limited than that relating to conceptual models in general. Several work has made attempts to improve goal model comprehensibility through involving users in the design process. Horkoff et al. [33] presented a framework purposed for interactive and iterative analysis. The process in their work include of interactively involving users when specifying goals in early requirement engineering through asking break-down questions, which they referred to as *Forward Analysis*. Their framework is based on using the *i** modeling framework, adopting the NFR satisfaction labels. Given the similarities between *i**, NFR [15], GRL [2], and Tropos [10], the authors deemed their framework applicable to these modeling methods, but remain unsure about other methods in the literature.

The way various concepts within goal models are visualized has also been the matter of investigation and empirical evaluation. Moody et al. [52] offer an assessment of the *i** visual syntax based on established rules ("Physics of Notations"- Section 2.3) addressing flaws in the modeling methods regarding its different visual elements, and suggesting probable improvements that can help increase understandability when perceived by end-users. An empirical analysis was followed by Caire et al. [12] in which experimental participants evaluate visualization choices of the language's primitives. The experiment was based on presenting the concept of *Semantic Transparency* which promote the idea of symbols visually representing its meaning to reduce the cognitive loads on the end-users side. Carvallo and Franch [13] have also studied, in the context of a case study, how non-technical stakeholders performed in developing strate-

gic dependency *i** diagrams, promoting through their study, the possibility to involve stakeholders in the process of context model construction.

Elsewhere, Hadar et al. [27] compare goal diagrams modelled using Tropos goal-oriented framework with those modelled using scenario-based framework Use Case diagrams on a variety of user tasks. The goal of the study was to compare which of models are more comprehensible to analysts. Measures include text-model mapping, model reading (extracting information from the model), and model modification (performing targeted modifications to models). The results have shown that, in cost of time, Tropos tend to exceed Use Cases in all three tasks.

Compared to the above efforts, the work presented in this thesis fits in a context of an on-going program to understand the comprehensibility and intuitiveness of a specific construct within goal models: contribution links. In earlier work, for example, Liaskos et al. [43] investigated the qualitative propagation rules of Table 2.1. Through an experiment of a nature similar to the one described here, it was observed, among other things, that positive labels and satisfaction values appear to be more readily understandable than negative labels and denial values.

Likewise, Alothman et al. [5] have also compared the various models for quantitative satisfaction propagation including the one used here and three versions of the one proposed by Giorgini et al. [26]. Their work simply presented to participants hierarchies of soft-goals with known satisfaction values at the leaf level, and asked them to choose the satisfaction of the root goal from a set of four values, each representing one of the possibilities of Table 2.2. It was found that the serial-parallel method was not preferred method while the most preferred depended on whether the contribution weights added up to 1.0, in which case a linear interpretation was evoked.

In a fashion somewhat similar to that of Caire et al. [12], Liaskos et al. [39] explored

visual ways for representing contribution levels instead of symbols and numbers and found that even simple combination of pie-graphs and bar-graphs allow for better accuracy. The difference this effort and the one presented in this thesis is that, while in that paper the semantics are assumed and the visualization is in question, in the study proposed here, both the visualization and its meaning are under comparison.

Chapter 3

Methodology And Experimental

Design

3.1   Overview

The motivation of this experiment presented in this chapter is to search for evidence that would offer some answers to the research questions presented in Chapter 1. Recall that these research questions are as follows: firstly, [RQ1] we are interested in learning which of the two methods of modeling contributions links is better in terms of usability in general, and in particular which of the two methods is more intuitive [RQ1.1] and more efficient [RQ1.2]. Then [RQ2] we want to investigate and compare the role of individual differences in affecting the comprehension and intuitiveness of the two modeling methods. Individual differences areas of interest being: [RQ2.1] cognitive styles, [RQ2.2] mathematical anxiety, [RQ2.3] arithmetic ability, and finally [RQ2.4] the method/approach followed by individual with solving the goal model exercises.

The experiment we conducted to address the above questions has a confirmatory and an exploratory aspect. It first aims to experimentally compare two approaches for modeling contribution in goal models, qualitative and quantitative. With regards to (RQ1), we specifically aim at identifying which of the two representation approaches is more intuitive and efficient allowing for more accurate and quick decision making. To this end, we posit the main following hypothesis:

Hypothesis 1 Goal Models using quantitative contribution link approach are more intuitive than those using the qualitative approach.

We further want to explore whether and how individual differences and ways of working, specifically ability and attitude towards math, cognitive style as well as followed approach, affect intuitiveness dimensions (RQ2). Given the absence of earlier experience in this particular aspect of the research, no explicit hypotheses are made with regards to (RQ2). Instead, implications are made through observations and study of the experiment results.

3.2 Experimental Design

3.2.1 Overview

In order to fulfill the objectives of the study, an experimental protocol was designed for collecting data. Specifically, a number of goal models of various sizes displaying decision problems in various domains have been developed. These models represent decisions with two or three alternatives and are specially crafted so that one of the alternatives is optimal with respect to the criteria structure modeled through a hierarchy of contributions. There are two groups of models: those modelled though quantitative

and those modelled through qualitative contributions.

Experimental participants are split in the two groups in a between-subjects fashion: some are exposed only to qualitative models and some only to quantitative. Each participant is presented a sequence of the experimental models and decide which of the alternatives represented in each model is the optimal one. The correctness of their aggregated (by group) responses marks the accuracy. Response time is also captured to be used as a measure of usability of the representation.

Prior to performing the above exercise, users perform a pre-test whereby psychometric characteristics are identified. The elicited psychometric and demographic characteristics are then used for exploring how they interact with accuracy and efficiency measures identified in the main test.

3.2.2   Constructs and Variables

The central construct of our experiment is *intuitiveness* as discussed in section 2.4. To measure it, we expose experimental participants to a set of models and ask them to perform inferences based on the information given in the model. The participants have only basic awareness of the modeling language and the abstract meaning of the constructs but no knowledge of precise semantics. Intuitiveness is measured primarily via *accuracy* of the participant inferences, i.e., the number of inferences that match the ones that the language semantics dictate. The more the matching ("correct") inferences the more the agreement between the user and the language designers, and, hence, the more intuitive the visualization-semantics package can be considered.

In addition to accuracy, we also measure *efficiency* when applicable, which is the number of accurate (matching) responses divided by the time it took to make the necessary inferences. There is also the construct of *confidence* levels of the method followed

by as the participants, as well as that of the inferences made and both are self reported by the participants themselves.

3.2.3 Instruments and Materials

The instrument used to acquire our data is a sequence of on-line screens containing directions, training materials, questions and tasks. It starts with a consent form that explains the purpose of the research and the nature of the tasks the participants are required to do. This form is followed by questions about demographic information such as age, sex, education and background in addition to their familiarity with goal models. In addition to this consent form, some training materials are included followed by a number of adapted and customized measures were implemented in the survey, and in what follows, details on each of these materials and measures:

Training

Participants are offered two video presentations introducing them to the concepts of decision alternatives and criteria, as well as goal models and the high-level meaning of either type (depending on instrument) of contribution links. Care is taken so that: (a) the videos are as much as possible identical to each other (e.g. use of same examples and points, about same length, same narrator, same visuals etc.), (b) the videos do not prescribe any exact method for interpreting satisfaction propagation that would allude to specific semantics.

Cognitive Styles Measure

Cognitive Style Index (CSI) by Allinson and Hayes (1996) [4] was adopted to serve the purpose of this research. The CSI is a 38-item self-report questionnaire[1]. Each item

---

[1]The exact instrument cannot be reproduced here due to copyright restrictions.

has 'true', 'uncertain' and 'false' response options, and scores of 2, 1 or 0 are assigned to each response with the direction of scoring depending on the polarity of the item. The nearer the total score to the maximum of 76, the more 'analytical' the respondent, and the nearer to the minimum of zero, the more 'intuitive' the respondent.

Math Anxiety Measure

Abbreviated Math Anxiety Scale (AMAS) by Hopko et al. [32], which is concerned with measuring the mathematics anxiety level of each individual. It is a 9-item self-report questionnaire that uses a five-level Likert type scale as shown in Figure 4. Subjects will rate the degree to which they agree or disagree with a statement from "strongly agree" as the highest degree, "agree", "undecided", "disagree" or "strongly disagree" as the lowest degree. The measure is presented in the survey as it appears in Figure 4.

Mathematical Ability Measure

As opposed to the above measures that are standardized, the Mathematical Ability Measures we employed are custom made for the needs of the experiment. It consists of a series of exercises in mental arithmetic, the list of these exercises are presented in Table 3.1 and 3.2 Subjects are expected to solve (28) mathematical problems with escalating complexity level. It consists of four (4) direct multiplications scored in $[0 \ldots 10]$ though an exponentially decaying function of the absolute distance between participant response and correct answer, four (4) comparisons of two-number products with a 0.05 distance and two (2) comparisons of two linear combinations containing two terms with a 0.25 distance. The overall score is normalized to the $[0 \ldots 12]$ for uniformity with the accuracy and other scores. The math exercises appear in the survey

Figure 4: Abbreviated Math Anxiety Scale (AMAS) by Hopko et al. [32]

as visualized in Figure 5 and 6.

Goal Model Comprehension Test

A number of goal models of different visualization were devised by the research investigators. These models of decision making come from three different domains: choosing a department, choosing an academic course, and choosing a mean of transportation. These domains were selected to represent a realistic setting to which the experiment subjects can related to. Two model structures were devised for each domain (a "small" and a "large"), and for each structure, two sets of labels of different values were devised, which result in 12 models. The 12 models exist in two copies: one with qualitative labels (seen in Figure 7, 8 and 9), and one with quantitative labels

Table 3.1: Direct Calculation Math Exercises

Addition	Multiplication	Division
0.76 + 0.19 =	0.28 x 0.27 =	0.46 / 0.62 =
0.47 + 0.73 =	0.95 x 0.49 =	0.44 / 0.54 =
0.77 + 0.28 =	0.15 x 0.34 =	0.37 / 0.88 =
0.89 + 0.66 =	0.74 x 0.45 =	0.18 / 0.73 =

Table 3.2: Comparison Math Exercises

Addition	Substraction	Multiplication	Division
0.92 + 0.16	0.66 - 0.46	0.51 x 0.41	0.54 / 0.73
0.25 + 0.89	0.81 - 0.66	0.32 x 0.82	0.69 / 0.88
0.36 + 0.72	0.48 - 0.11	0.48 x 0.29	0.25 / 0.98
0.59 + 0.54	0.61 - 0.29	0.12 x 0.79	0.11 / 0.55
0.55 + 0.36	0.75 - 0.35	0.03 x 0.71	0.52 / 0.72
0.72 + 0.24	0.42 - 0.06	0.37 x 0.19	0.59 / 0.87
0.85 + 0.51	0.81 - 0.12	0.19 x 0.83	0.54 / 0.73
0.37 + 0.95	0.93 - 0.28	0.32 x 0.36	0.69 / 0.88

(seen in Figure 10, 11 and 12), all else being equal. Hence, the total models produced are 24. Each goal model presented in the experiment as a graphical image representing the domain of the exercise. By the end of the exercises, participants are expected to self-report the *Approach* they followed, between "using their intuition" and "following a specific method".

The labels of the contribution links of the models follow a specific random sampling technique, which we describe below:

- Contributions of quantitative models are a random sample from possible results of AHP comparison processes, such that the score distance between the first and the second option does not exceed 0.4 (nearly half of the maximum distance), according to the Liaskos et al interpretation [40].

- Contribution of qualitative models are again randomly sampled from the domain of four contribution labels ("–", "-", "+", "++") in a way that the first and second option do not exceed half of the maximum possible distance, according to

Figure 5: An 'Addition' Direct Math Exercise

Figure 6: A 'Substraction' Comparison Math Exercise

Giorgini et al. [25].

Quantitative Goal Model Example

For clarity on how the models are constructed and how they encode a decision problem, we go over the calculations based on the goal model of Figure 13. For the quantitative goal model exercise about the Apartment Choice depicted in Figure 13, we follow the linear method proposed in 2.2 to find the satisfaction value of each choice individually, and then compare the end value of each result to find the optimal answer. As observed in Figure 13, we have two choices (Apartment 1, Apartment 2) to choose from, meaning the calculations take two stages.
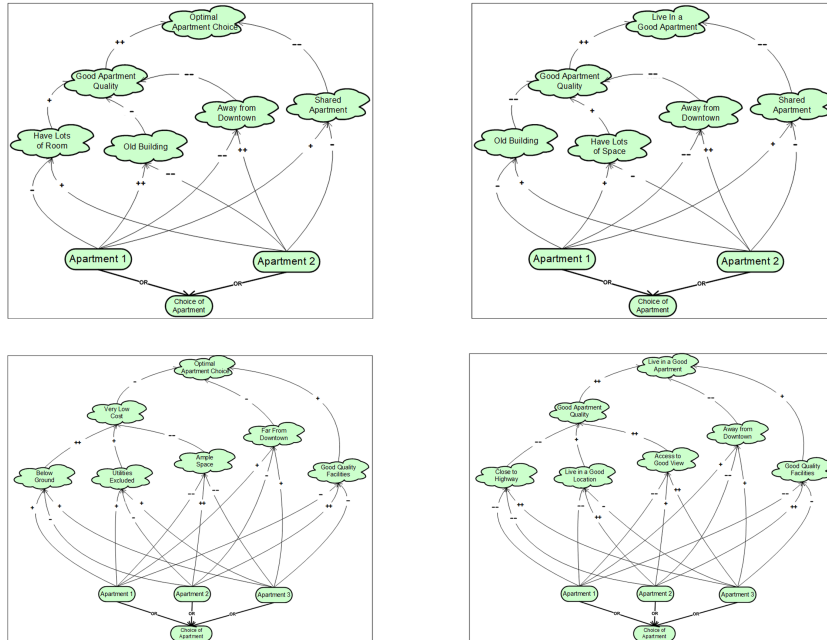
Figure 7: Experiment Goal Models with Qualitative Labels (Apartment)
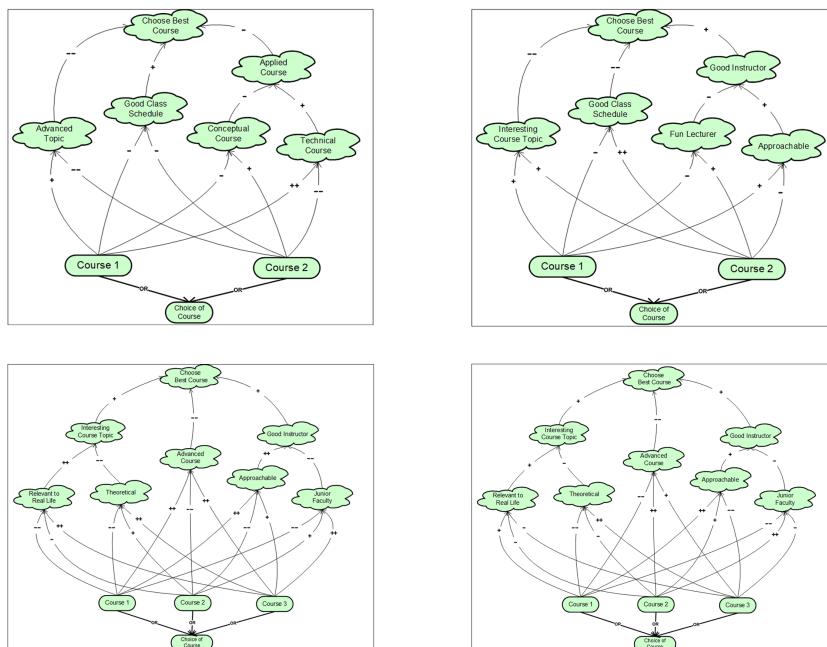
Figure 8: Experiment Goal Models with Qualitative Labels (Course)

Figure 9: Experiment Goal Models with Qualitative Labels (Transportation)

Figure 10: Experiment Goal Models with Quantitative Labels (Apartment)
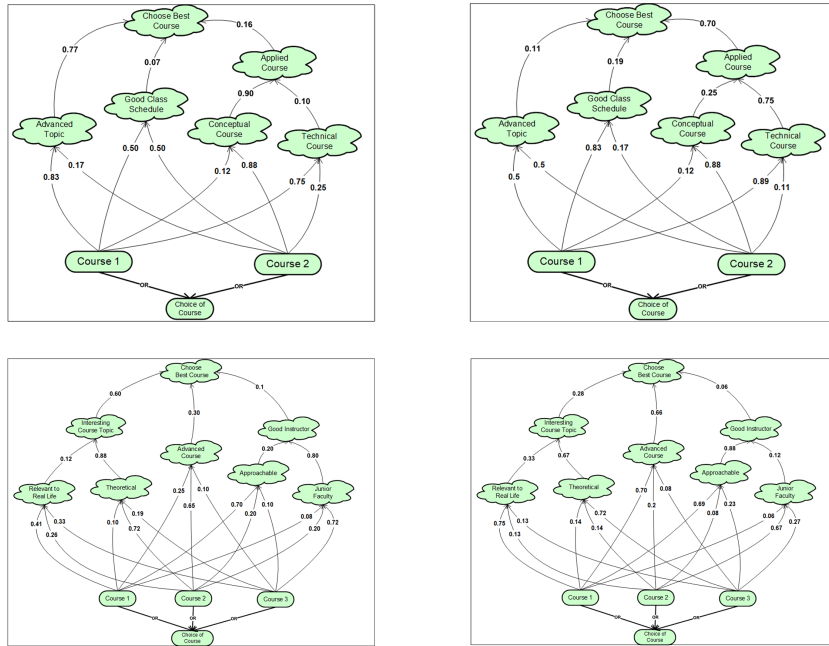
Figure 11: Experiment Goal Models with Quantitative Labels (Course)
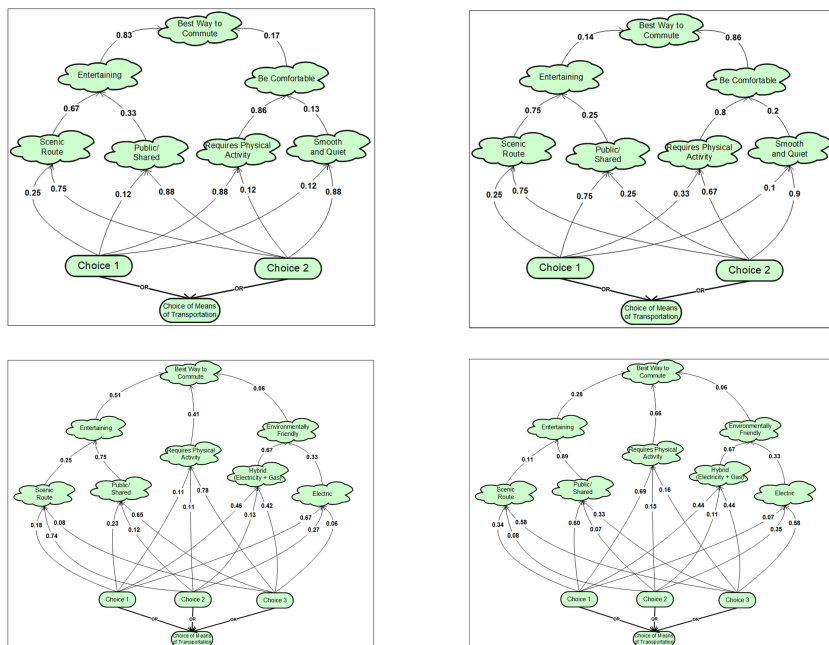
Figure 12: Experiment Goal Models with Quantitative Labels (Transportation)

(a) Qualitative Goal Model Exercise (b) Quantitative Goal Model Exercise
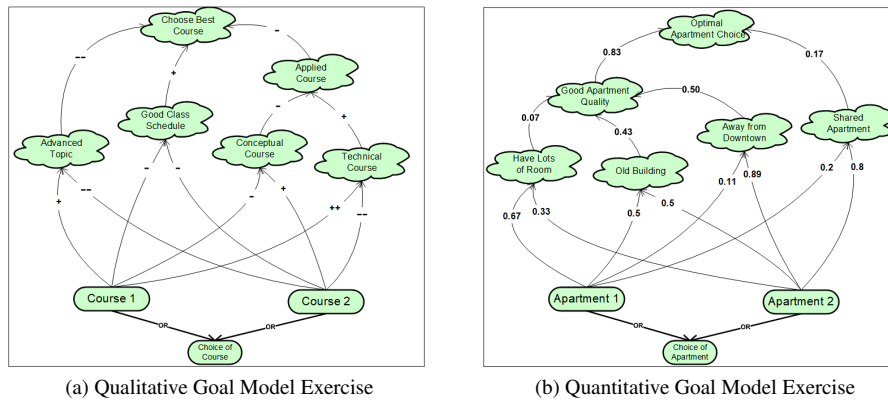
Figure 13: Survey Goal Model Exercises

For stage one in which we consider Apartment 1, let the base satisfaction value of Apartment 1 be 1.0, and the base satisfaction value of Apartment 2 be 0.0. We start the calculations by multiplying the satisfaction value of each option by the value of its contribution to the satisfaction of upper goals, starting with 'Have Lots of Room' of Apartment 1 as shown below:

$$1.0 \times 0.67 = 0.67$$

We repeat the multiplication for Apartment 2 as following:

$$0.0 \times 0.33 = 0.0$$

Next, we sum the result of both multiplication which will give us the satisfaction value of 'Have Lots of Room' of the first Apartment one:

$$0.67 + 0.0 = 0.67$$

Following the same calculation for the next goal 'Old Building' results in:

$$(1.0 \times 0.5) + (0.0 \times 0.5) = 0.5 + 0.0 = 0.50$$

We repeat the same process again for 'Away From Downtown':

$$(1.0 \times 0.11) + (0.0 \times 0.89) = 0.11$$

To calculate the satisfaction value of 'Good Apartment Quality', we first repeat the multiplication process of each of the lower goals founded satisfaction value by its con-

49

tribution to the quality of the apartment as following:

$$0.67 \times 0.07 = 0.05$$

$$0.50 \times 0.43 = 0.22$$

$$0.11 \times 0.50 = 0.06$$

Just like previous step, we then sum the results to give the satisfaction value of 'Good Apartment Quality':

$$0.05 + 0.22 + 0.06 = 0.33$$

Repeating the same steps for calculating 'Shared Apartment' satisfaction value:

$$(1.0 \times 0.2) + (0.0 \times 0.8) = 0.2$$

Finally, we calculate the satisfaction value of Optimal Apartment Choice of Apartment 1 by again multiplying the values of 'Good Apartment Quality' and 'Shared Apartment' by their contribution and sum the results as follows:

$$(0.33 \times 0.83) + (0.2 \times 0.17) = 0.27 + 0.03 = 0.30$$

So, the satisfaction value of Apartment 1 as an Optimal Apartment Choice is 0.3

Stage two of this exercise, we consider Apartment 2 by assuming its base value to be 1.0 and the base value of Apartment 1 to be 0.0, and by repeating the steps we followed in Stage one, the calculations are as following:

$$\{(0.0 \times 0.67) + (1.0 \times 0.33)\} \times 0.07 = 0.02$$

$$\{(0.0 \times 0.5) + (1.0x0.5)\} \times 0.43 = 0.21$$

$$\{(0.0 \times 0.11) + (1.0 \times 0.89)\} \times 0.50 = 0.45$$

$$\{(0.02 + 0.21 + 0.45) \times 0.83\} + \{((0.0 \times 0.2) + (1.0 \times 0.8)) \times 0.17\} = 0.7$$

This gives us the overall satisfaction value of Apartment 2 as an Optimal Apartment Choice to be 0.7, and by comparing the calculations, we conclude with the fact that Apartment 2 is the better choice over Apartment 1 which has an overall satisfaction value of 0.3. We further observe that the distance between the two scores is 0.7-0.3

= 0.4, which satisfies the distance requirement between first and second option (see below).

Qualitative Goal Model Example

For the qualitative goal model exercise, we take the Course Choice qualitative example depicted in Figure 13. We follow the method proposed by Giorgini in [25] to find both the satisfaction and denial values of each choice individually, and then compare the end value of each choice to find the optimal answer. As we can see in the example figure, we also have two choices (Course 1, Course 2) to choose from, which also means the calculations take two stages. The difference for the qualitative method is that instead of math calculations, we follow the rules of Table 2.1 in Chapter 2 to decide the satisfaction and denial values of a goal based on the effect of the contribution label.

However, to facilitate description of the algorithms without delving into the complexity of the Giorgini et al. formalisms, we first need to define an evidence maximization function as follows: Let L be a set of candidate satisfaction and denial values for a goal, that are the result of applying propagation rules to all goals that have incoming contribution values to the goal. For example, $L = \{FD, PD, PS, N, FD\}$; each of these elements are the result of combining each of the satisfaction and denial values of the origin with the contribution label.

Then $Infer(L)$ returns a pair of values $<,>$ each representing the highest satisfaction and highest denial evidence found in L, respectively. In the example above $Infer(L) = Infer(\{FD, PD, PS, N, FD\}) = < PS, FD >$, because the highest satisfaction label is PS (the third element) and the highest denial label is FD (the first and last elements).

For stage one in which we consider Course 1, let the base satisfaction value of

Course 1 be Fully Satisfied (*FS*), and the base satisfaction value of Apartment 2 be None (*N*). Following the rules in Table 2.1, we find the satisfaction value of 'Advanced Topic' for Course 1 to be:

$$FS \xrightarrow{+} PS$$

And for Course 2 to be:

$$N \xrightarrow{--} N$$

Next, we maximize both results of Course 1 and Course 2, which will give us the satisfaction Value of 'Advanced Topic' to be:

$$Infer(\{PS, N\}) = < PS, N >$$

Following the same process for the next goals of 'Good Class Schedule', Conceptual Course' and 'Technical Course' results in:

Good Class Schedule: $Infer\{(FD(-), N(-))\} = Infer\{(PS, N)\} = < PS, N >$

Conceptual Course: $Infer\{(FS(-), N(+)\} = Infer\{(PD, N)\} = < N, PD >$

Technical Course: $Infer\{(FS(++), N(--)\} = Infer\{(FS, N)\} = < FS, N >$

To find the satisfaction value of 'Applied Course' we calculate the values of both Conceptual Course and Technical Course, then maximize the result as the following:

Applied Course $= Infer\{(PD(-), FS(+)\} = Infer\{(PS, PS)\} = < PS, N >$

Finally, we calculate the satisfaction value of 'Choose Best Course' of course 1 by again calculating the values of 'Advanced Topic' and 'Good Class Schedule' and 'Applied Course' with their contribution links and then maximizing the results:

$$PS \xrightarrow{--} PD$$

$$PS \xrightarrow{+} PS$$

$$PS \xrightarrow{-} PD$$

$$Infer\{(PD, N, PS, N, N, PD)\} = < PS, PD >$$

So, the satisfaction value of Apartment 1 as an Optimal Apartment Choice is $< PS, PD >$

For stage two, we consider Course 2 by assuming its base value to be Fully Satisfied (*FS*) and the base value of Course 1 to be None (N), and by repeating the steps we followed in Stage one, the calculations are as following:

Advanced Topic: $Infer\{(N(+), FS(--)\} = Infer\{(N, FD)\} = <N, FD>$

Good Class Schedule: $Infer\{(N(-), FS(-)\} = Infer\{(N, PD)\} = <N, PD>$

Conceptual Course: $Infer\{(N(-), FS(+)\} = Infer\{(N, PS)\} = <PS, N>$

Technical Course: $Infer\{(N(++), FS(--)\} = Infer\{(N, FD)\} = <N, FD>$

Applied Course: $Infer(PS(-), FD(+)) = Infer\{(PD, PD)\} = <N, PD>$

Choose Best Course: $Infer\{FD(--), PD(+), PD(-)\} = Infer\{(FS, PD, PS)\}$

So, the satisfaction value of Course 2 as an Optimal Apartment Choice is $<FS, PD>$

We conclude that Course 2 is the best course choice with the satisfaction value of $<FS, PD>$ as opposed to Course 1 which has the satisfaction value off $<PS, PD>$.

Model Sampling

As we saw, we develop the goal models used for the experimental instrument by picking a goal structure and populating the contribution links with random contribution labels such that the optimal alternative has a fixed distance from the second optimal one, as measured by the satisfaction each induces to the root goal. We require this restriction so that there is sufficient difference between the best and second best to allow for some intuitive detection, but the difference is not too much to be too obvious for all participants.

Calculating the distance from best to second best alternative is straightforward in the case of numeric models. The choice of each alternative will result in a number representing the satisfaction value of the root soft-goal for that alternative. We simply ensure that the largest value is about 0.4 higher than the second best. For the symbolic

models the comparison is less straightforward due to the presence of both a satisfaction and denial value. To be able to perform a comparison, we aggregate the two values into one. To do so we firstly associate qualitative satisfaction labels {N, P, F} with numeric values 0,1,2, respectively. Let then $sat(g)$ and $den(g)$ be the resulting numeric satisfaction and denial values for goal g. The aggregated satisfaction value is then $sat(g) - den(g)$ which is an integer in [-2,2]. For example, the aggregated satisfaction value of a goal g_1 with {**PS**, **FD**} is $sat(g_1) - den(g_1) = 1 - 2 = -1$ and of a goal g_2 with {**FS**, **ND**}, $sat(g_2) - den(g_2) = 2 - 0 = 2$. Given this aggregation procedure, we demand that our sample models have a distance of 2 satisfaction levels. For example, a label configuration in which the best alternative makes the root goal {**FS**, **ND**}, hence aggregated value $2 - 0 = 2$ and the second best makes the root goal {**PS**, **PD**}, hence aggregated value $1 - 1 = 0$ passes the qualification qualifies. In the example of the previous section (Figure 9) the first option was $\{FS, PD\}$ and the second $\{PS, PD\}$, so the total distance is 1 - 0 = 1, which is too small a distance to pass the test. To see why this distance matches the one chosen for the symbolic models to allow for a fair comparison, observe first that the maximum distance between alternatives in the symbolic case in terms of aggregated value is 4 ({FS, ND} versus {NS, FD} so $2 - (-2)$). The distance we demanded in symbolic models is 2, thus half of the space. Observing now that the corresponding maximum distance in numeric models is 1.0, it follows that half-space size distance would be 0.5. However we end up with 0.4, slightly biasing against numeric models, as for some of our structures we fail to find label configurations that yield 0.5 distance.

3.2.4   Procedure and Administration

The experiment procedure has two phases that are described in what follows:

Phase One

Phase one starts with the pre-test in which each participant responds to the CSI 38-item questionnaire. Next, they answer the AMAS 9-item questionnaire about their anxiety level with mathematics. After a little break, they are exposed to the 28 arithmetic exercises to solve, starting with addition exercises, followed by multiplication and division. The last part of the pre-test is a number of quantitative comparison questions, in which participants compare between two equations. During the mathematics exercises, they are instructed not to take notes, get help or use any calculator devices. This phase ends with a question about their sampling identity whether they were university students or Mechanical Turks, and then they are provided with a participation code.

Phase Two

In the beginning of phase two, participants start with entering the participation code they were provided in the pre-test, which purpose is to link the two answered tests to the same participant. Participants then answer questions about their demographic information and their background knowledge in goal models.

Subsequently, participants are sequentially presented with the goal models, as shown in Figure 13, and are asked to enter which of the two or three alternatives they think is optimal. In the end, they are asked if they used a specific method in making their decision, and what that method is, or whether they used their intuition.

We note that midway in the data collection process, the instrument underwent the following revisions: (a) the math ability test was changed and moved to the end and (b) two additional questions asking for the participant's confidence in their responses and also the method they use method were added. Note also that although the intent was that the first and second phases were meant to be administered at separate times,

student participants were given the choice to participate in one session, which they all did.

3.2.5 Participant Sampling

The sampling methods for the proposed experiment are both non-probability voluntary and random sampling. Specifically, participation is sought from two sources:

(a) Third year undergraduate students of the School of Information Technology at York University.

(b) Mechanical Turk (MT) participants with a US college degree.

We consider both populations suitable for two reasons: (a) the tasks appeal to simple cognitive functions and not domain specific, IT-related knowledge or experience, and (b) we believe goal models should be a tool for decision exploration, employable for any stakeholder involved in a project including non-technical ones (clients, users, owners etc.). Administration in the student population took place in the lab and allows for reliable response time measurement. Due to the absence of such conditions, for MT participants reliable response times are not assumed. In addition, as earlier student results tended to skew towards more analytical CSI scores, MT participant sampling was biased towards a more balanced CSI scores. This was achieved by separating the CSI instrument as a pre-screening test and then proceeding with participation invitations from the most intuitive respondents to the more analytical ones.

Chapter 4

Results and Analysis

4.1 Data Descriptives

4.1.1 Demographics

A total of 102 participants are included in the analysis: 27 students (21 males and 6 females) and 75 MT participants (41 males and 34 females). Table 4.1, 4.2 and 4.3 show the participants distribution based on sex, education and background. A few of the collected responses are excluded on the basis of not properly attending the training videos, as established by the limited time spend in the appropriate screens. The distribution of subjects across the two conditions of numeric and symbolic goal models contribution links is shown in Table 4.4.

Table 4.1: The Number of Participants by Sex

		MTurk	**Student**
Sex	Female	34	6
	Male	41	21

Table 4.2: The Number of Participants by Education

		Highschool	**Masters**	**Other**	**PhD**	**Post-Secondary**
Sex	Female	6	3	0	3	28
	Male	15	6	3	1	37

Table 4.3: The Number of Participants by Background

		Art	**Business/Econ**	**Education**	**Health**	**Humanities**	**Other**	**Social**	**STEM**
Sex	Female	3	7	2	1	4	0	7	16
	Male	3	15	1	2	4	1	3	33

Table 4.4: The Number of Participants by Condition

			Group	
			Numeric	**Symbolic**
Sample	MTurk	Female	20	14
		Male	18	23
	Students	Female	4	2
		Male	10	11

4.2 Individual Difference Metrics

4.2.1 Cognitive Style Index (CSI)

Based on the collected results, the overall CSI average was 47.91 which is above re-ported averages in the literature (about 44.53 according to the CSI manual [4] and Hmieleski & Corbett (2006) studying US college students [31]). Table 4.5 shows the specific average per sample type. In addition, the number of cases per High versus Low
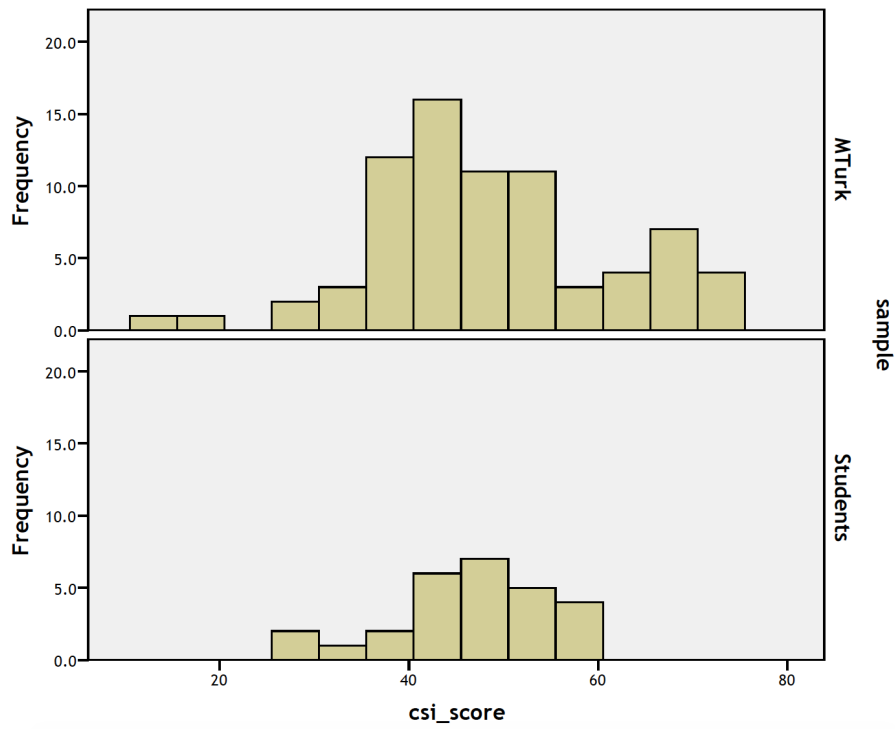
Figure 14: Participants CSI Score Frequency

CSI level based on the CSI Index with regards to population average is shown in Table 4.6, while Table 4.7 shows the CSI level frequency based on the sample type and the condition groups. A depiction of the CSI score frequency is presented in Figure 14.

Table 4.5: CSI Score Distribution

	Sample		
	MTurk	**Student**	**Total**
csi_score	48.43	46.48	47.91

Table 4.6: CSI Level Distribution

CSI Level	
High	Low
61	41

In general the Mechanical Turk sample appeared to be more "Analytical" (higher CSI index) than the Student sample. Despite the efforts for more participation of intuitive types, the averages are both higher than the normative average mentioned above.

Table 4.7: High And Low CSI Score Frequency

				CSI Level	
				High	Low
Sample	MTurk	Group	Numeric	21	17
			Symbolic	24	13
	Students	Group	Numeric	7	7
			Symbolic	9	4

4.2.2 AMAS

Based on the collected responses, the overall AMAS average is 20.28 which is just a bit below the reported averages in the literature (about 21.1 according to D.R. Hopko et al. 2003 [32]). The score is lower among graduates than it is among current students. Table 4.8 shows the specific samples and their average AMAS score (High Vs. Low). AMAS level frequency is shown in Table 4.9, and depicted in Figure 15.

Table 4.8: AMAS Score Distribution

	Sample		
	MTurk	Students	Total
amas_score	19.88	12.41	20.28
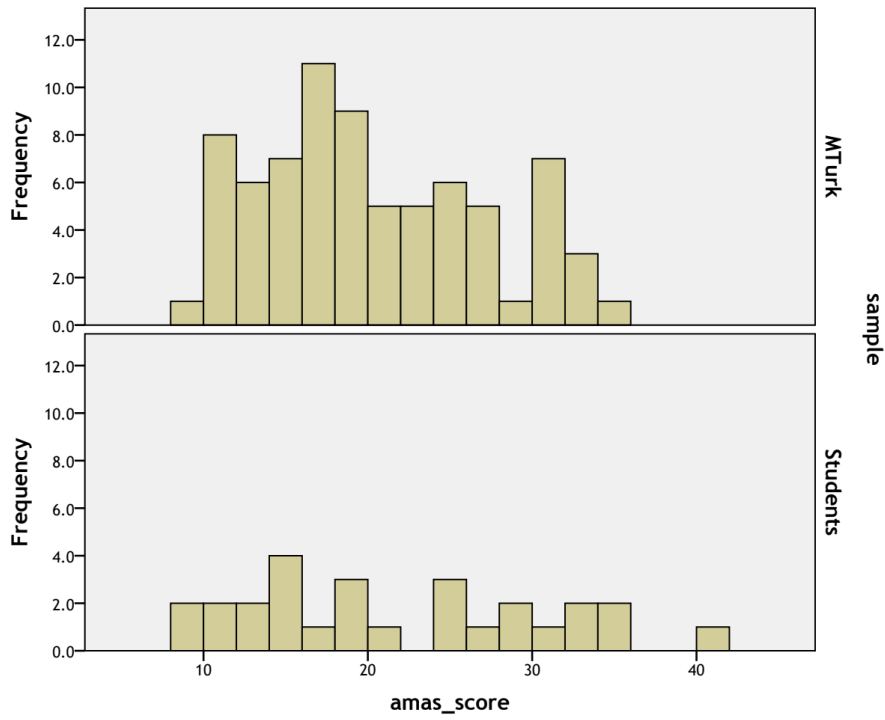
Figure 15: Participants AMAS Score Frequency

Table 4.9: AMAS Level Distribution

AMAS Level	
High	Low
44	58
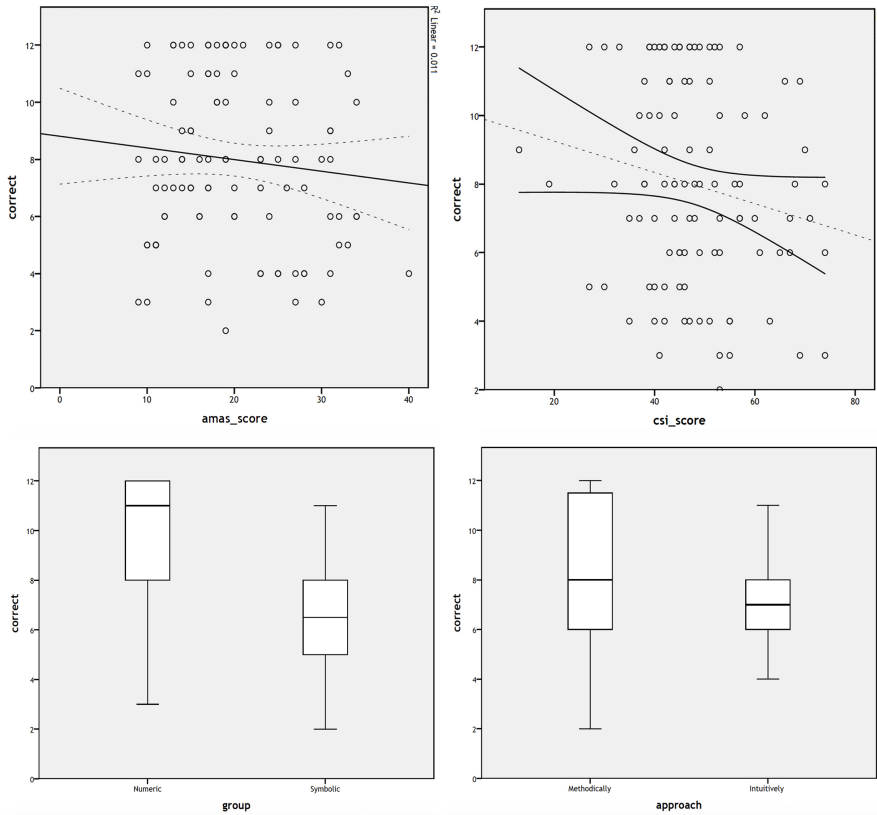
Figure 16: Accuracy Analysis

4.3 Statistical Analysis[1]

4.3.1 Accuracy Analysis

Accuracy is measured as the raw number (out of 12) of correct (wrt. semantics) choices of optimal alternative. To assess accuracy we first attempt to fit a linear model [70] including Representation, AMAS score, CSI score, and Approach as main effects, ignoring interactions for the moment. In other words we conduct analysis of variance with factors Group (qualitative, quantitative) and Approach (methodically, intuitively), while CSI scores and AMAS scores were included as covariates.

---

[1] The results of this experiment have been previously publish on [44], and more details on the statistical analysis can be found here http://tinyurl.com/y3koz33a
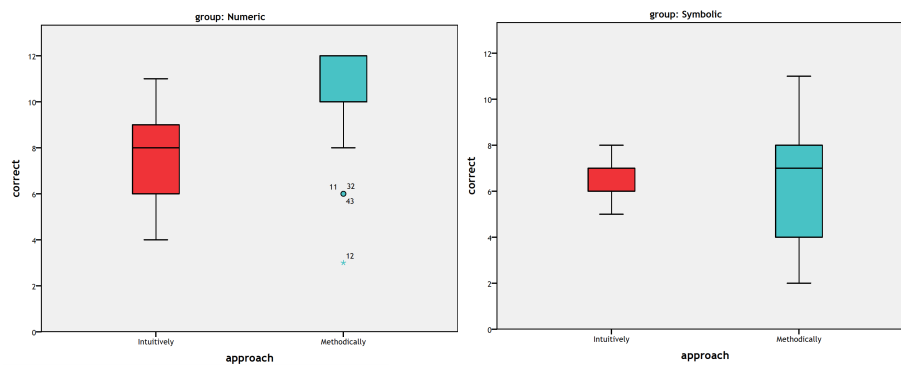
Figure 17: Accuracy Analysis Per Representation and Approach

Looking at the graphical result, represented in Figure 16, and the grand means of the two groups, we can see that the Numeric Group performed significantly better (75.7% correct) relative to the Symbolic Group (52%; $F(1,97) = 72.18$, $p < 0.001$, $d = 1.51$). Furthermore, when observing the effect of the working Approach, the methodical group performed better relative to intuitive group ($F(1,97) = 5.56$, $p < 0.05$, $d = 0.4$). The next observation with regards to Math Anxiety, the results reported a significant main effect of AMAS score ($F(1,97) = 5.65$, $p < 0.05$, $d = 0.33$). Thus, those with below average AMAS level (less anxious) score 0.96 more correct questions than those above average (higher anxiety level).

Considering a model with interactions we observe that the Working Approach variable interacts both with Representation and potentially with CSI score as seen in Figure 17. Specifically, we observe an interaction of Group and Approach ($F(1,91) = 6.66$, $p < 0.05$, $d = 1.37$). Performing a simple effects following [22] in which we fix the level in one factor and test for effects using the other factor, it was revealed that the main effect of group was significant in those participants who followed the methodical approach (numeric = 87.4%, symbolic = 51%; $F(1,77) = 77.34$, $p < 0.001$), but not in those who used intuitive approach (numeric = 61.5%, symbolic = 53.3%;
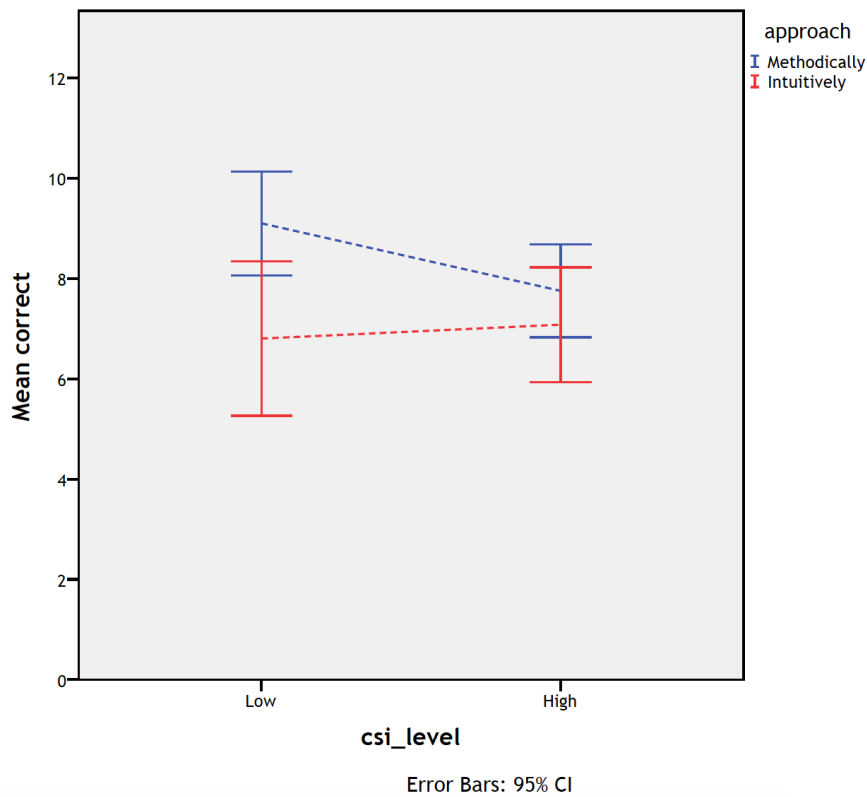
Figure 18: Accuracy Analysis Per Approach and CSI Level

$F(1,21) = 1.45, p > 0.2$).

An additional smaller effect, seen in Figure 18, is that those with more intuitive cognitive style have more to gain (1.6 out of 12 more correct answers) when working methodically instead of intuitively compared to the analytically inclined.

4.3.2 Efficiency Analysis

Efficiency, operationalized as the ratio of accuracy over total response time, is considered only for the 27 students of the corresponding samples, where response time can be reliably measured. As above we fit a linear model that involves Group (qualitative, quantitative) as a discrete factor, while the CSI scores and AMAS scores were included
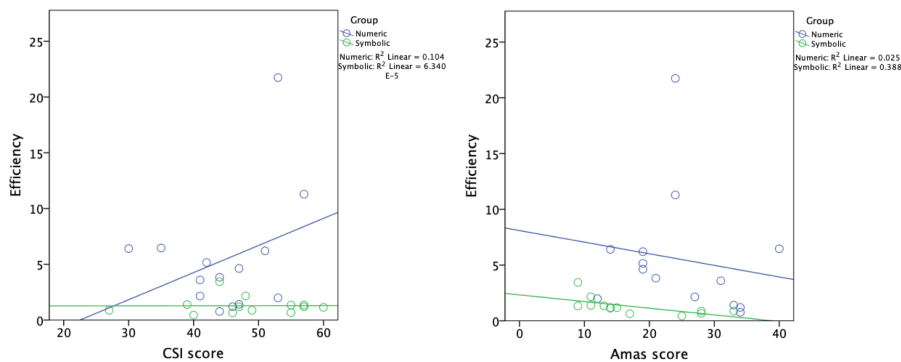
Figure 19: Efficiency Analysis Per CSI and AMAS Score

as covariates (Approach is not considered due to it being highly unbalanced).

We observe through the graphical result represented in Figure 19 a main effect of the Representation factor. The results show that the Numeric group produced higher number of correct responses per minute (0.54) relative to Symbolic group ($p < 0.01$); (0.12; $F(1,25) = 7.53$, $p < 0.05$, $\eta_p^2 = 0.259$). Due to a small sample size (N = 27) and the consequent possibility of non-normal distribution, we have additionally confirmed this result by means of a non-parametric Yuen's t -test, which also showed a significant difference ($t(9.41) = 3.77$, $p < 0.01$, effect size= -0.93). Finally, the assumption that either CSI or AMAS scores associates with efficiency is not supported by the data.

4.3.3 Confidence Analysis

Confidence measurement was introduced to the instrument for the last 45 MT participants only, and the analysis is based on that sample. It is measured on a 7-point "Likert"-style scale and treated as ratio as per common practice [61].

We again perform ANOVA test studying the main effects of representation, CSI level, AMAS level and math ability as well as interactions between them. The observation of the result shows an interaction of Group and AMAS level ($F(1,40) = 4.22$,
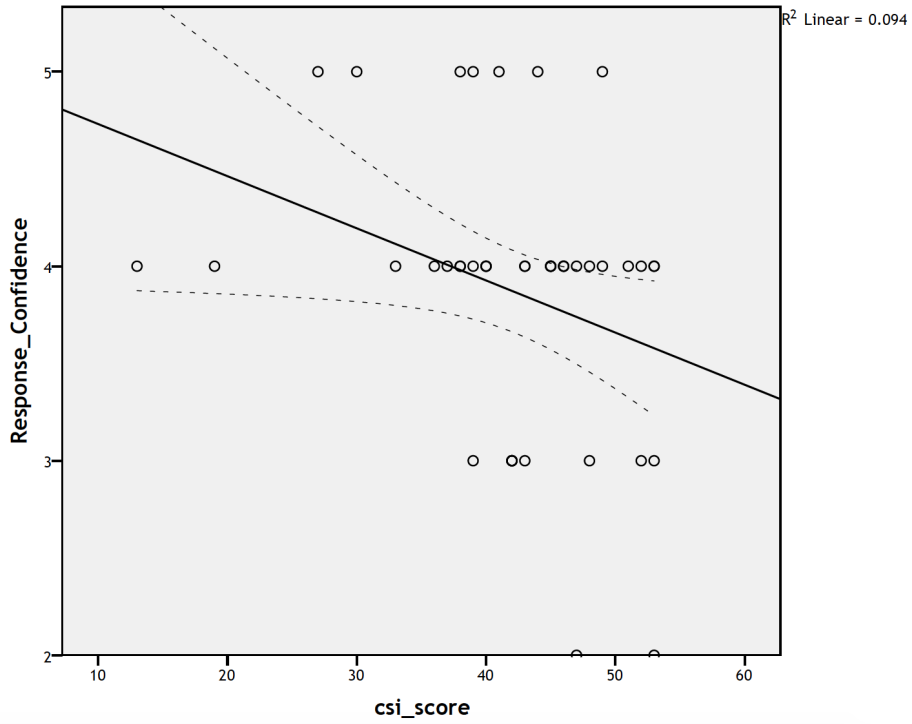
Figure 20: Confidence Analysis Per CSI Score
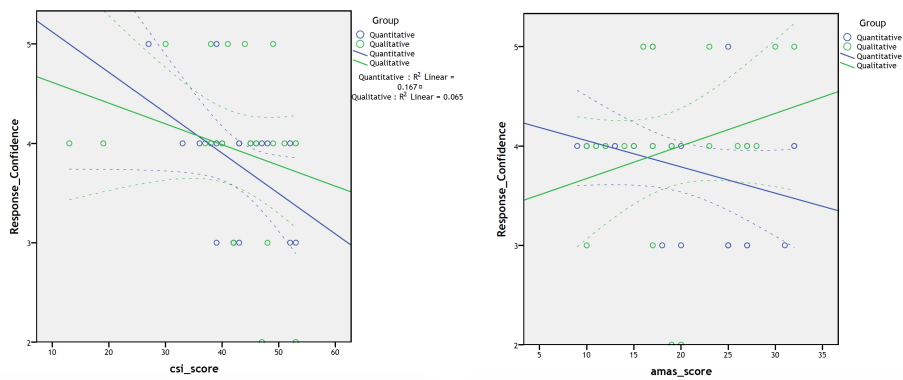
Figure 21: Confidence Analysis Per CSI Score and AMAS

$p < 0.05$, $d = -0.039$). Performing a simple-effects follow up [22], the result shown in Figure 21 depicts a main effect of Group on Response Confidence which was significant in the high AMAS level (qualitative = 4.43, quantitative = 3.56; $F(1,15) = 7.08$, $p < 0.05$, $\eta_p^2 = 0.336$), but not in the low amas level (qualitative = 3.73, quantitative = 3.93; $F(1,28) = 0.54$, $p > 0.4$, $\eta_p^2 = 0.02$). This observation shows that respondents with higher math anxiety level in the numeric group seem to have less confidence in their answer. The same procedure was performed with the Method Confidence as a dependent variable. However, this resulted in no significant main effect and/or interactions.
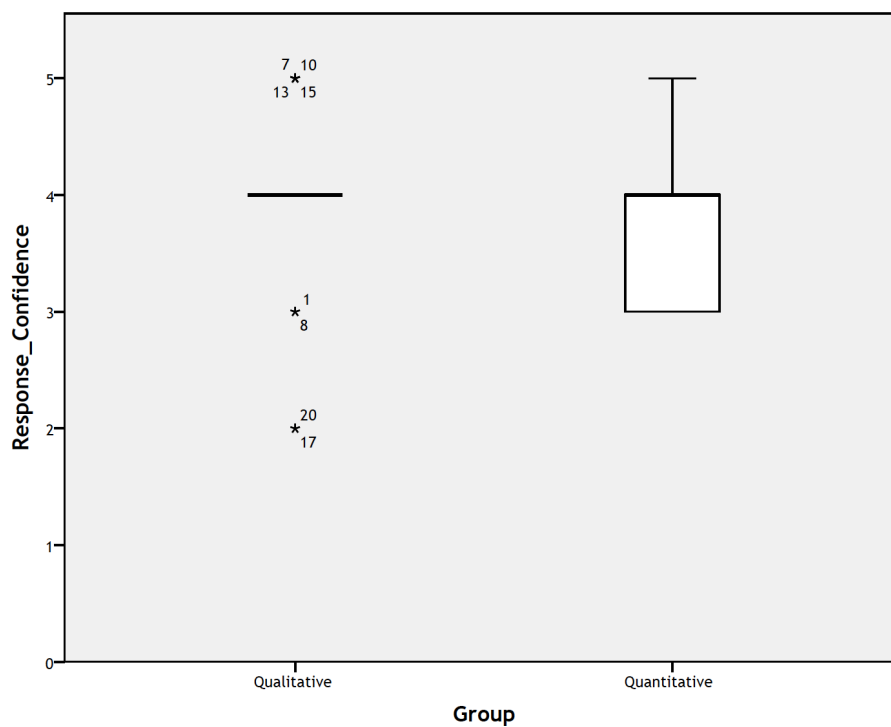
Figure 22: Confidence Analysis Per Representation

4.4   Summary and Explanatory Remarks

The results present substantial evidence that the numeric representation according to the linear model in Table 2.2 leads to more compliant decision-making inferences by untrained users and faster than the qualitative one of Table 2.2. We can attribute this to the familiarity that users have with numbers and proportions, on which the numeric model is based, and the lack thereof for symbolic labels. The effect emerges (strongly) when the participants say they work methodically, meaning that some explicit mental model is developed. Interestingly, however, representation does not influence choice of approach, meaning that a mental model is evoked for both kinds of models. However, in the symbolic case either the evoked model is in strong agreement with the authoritative one, or the latter is correctly guessed but poorly executed. At the same time, the general lack of correlation between arithmetic ability and accuracy, assuming that our instruments have any reliability, may indicate that participants in the numeric group may not perform mental calculations as per the linear equation in Table 2.2 (which would lead to errors) but base their success on an evoked heuristic that works as well. Furthermore, counter to our expectation that AMAS Level would affect only the numeric group it seems to affect both groups, implying the possibility that the requirement for either kind of symbolic inference is akin to a mathematical task, in which, in turn, highly math-anxious individuals turn to perform worse.

Indeed, focusing on 3-option models the responses of more than 82% out of those who respond "Intuitively" is not distinguished from random (using $p < 0.05$ binomial tests), compared to 55% out of those who responded "Methodically". Thus, self-reported intuitively might as well be construed as "randomly" or expressing inability to make a guess.

4.5 Validity Threats

We briefly address the most important of construct, internal, external and statistical conclusion validity.

4.5.1 Construct Validity

Construct validity is defined as the degree to which a test measures what it claims, or purports, to be measuring [17]. In other words, it is used to determine how well a test measures what it is supposed to measure. With regards to the measuring variables of the Intuitiveness construct in our work, our fundamental assumption that it can be measured by the alignment of spontaneous with authoritative inferences can be criticised as being too "particular" (i.e., addressing a very restricted kind and use of the model) and avoiding of examination of what goes on in participants' minds when confronted with an unknown notation. A possible response is pragmatic: the observed substantial effect on representation accuracy and efficiency is immediately usable even when theoretical clarity is pending: numbers seem to "just" be more intuitive for the particular task. However, careful follow-up qualitative investigation on the exact understanding technique that participants adopt for their reasoning is part of our research agenda. A further criticism can be extended to the ad-hoc development of non-standard math ability tests. The rationale behind this decision being took the absence of suitable standard instruments that deals with decimal numbers. However, the results seem to show not major effects regardless.

4.5.2 Internal Validity

Internal Validity is defined the extent to which a piece of evidence supports a claim about cause and effect, within the context of a particular study [59], i.e whether there

are other explanations that are rival to the ones we claim for the observed differences. Two main threats to internal validity in our work revolve around the representation factor. On one hand, the "difficulty" of the symbolic models (distance between first and second optimal) is constructed based on an operation of comparing satisfaction and denial values that may be argued to be arbitrary and off-specification (according to [25]). However, in our view, insofar as the two representations can be used for the same purpose (comparing alternatives) they cannot be considered incomparable. The subsequent question of what ways, other than the ones adopted here, can be consider for fairly constructing absolute preferability distance between satisfaction levels in a two-valued setting appears to be difficult to address. Furthermore, difference in training quality can be argued to work against one of the conditions. Such bias is difficult to measure and control for. We are hoping that our carefully scripted, video-recorded training videos (versus live lectures) offer a first line of defence against this threat.

4.5.3 External Validity

External validity is defined as how well the outcome of a study can be expected to apply to other settings outside that the study context. Threats to external validity with regards to generalizability in this particular study concentrate on the choice of participants, models, as well as the tasks the participants performed. We first claim that our participants being non-experts and (some of them) students does not harm generalizability. On one hand, there seems to be an implicit desire in the goal modelling community that non-technical stakeholders (users, owners, clients) should be able to use such models. On the other hand, although we could not find research that describes the typical characteristics of either business and systems analysts and their clients, we cannot just assume that there are exclusively of a technical background. In cases where

potential non-expert clients are involved by the analysts in the design process of an information system, they can exposed to the conceptual model of its design for for transparency purposes. Faced with these models, they may have to make decisions with regards to alternatives of both goals and functions of the systems, which is similar to what the exercises employed in our experiment. We, thus, find that our participants constitute a good sample of the population that may be a user of such a model. Furthermore, the choice of models that we used for the instruments brings unavoidable structural, size and domain commitments. Therefore, until research with different models is conducted, generalizations should be carefully done for models of similar characteristics. Finally, a significant external validity question is whether the effects identified in this study would be found in different tasks of similar nature that take place in real world use of the models, assuming otherwise similar goal models and participants. Generalization potential would then depend on how different the tasks are from the ones studied here. For instance, one can hypothesize that in numeric models, participants might be more accurate in describing how a specific goal alternative affects high-level goals. On the other hand, if the goal models are the result of eliciting user's beliefs about the contribution structure, the result does not say anything with regards to whether the user's choice is consistent with their input in developing the model; this is a completely different task. Thus, different proposed uses of goal models will require separate examination.

As a final note of *statistical conclusion* – which is concerned with the extent to which the conclusions drawn from a statistical test are accurate and reliable, while we pre-hypothesized the effect of representation format, the rest of the factors and interactions thereof where the result of some statistical model exploration in order to find well suited models. This research process took place in the spirit of informing future re-

search and confirmatory studies. Thus, despite reported statistical significance levels, except for the effect of representation, the remaining effects continue to be tentative and subject for further confirmation.

Chapter 5

Conclusion And Future Work

5.1 Conclusion

In this thesis, we designed, executed and analyzed the results of an experimental study which compared two ways of representing contribution links within goal models with regards to their intuitiveness, our main comparison construct. We also explored the effect of individual differences to the demonstration of intuitiveness. The main contributions of this work are listed in the following:

- Firstly, we offer the first, to our knowledge, empirical study of the two approaches of goal model contribution links mentioned above through the measuring of intuitiveness and efficiency. In the results, which addresses the concern of the research question [RQ1], we find that the models with numeric contribution links lead participants to more accurate responses when the latter are the result of adopting a specific method by the participants. This we believe immediately informs practice and motivates more research on the subject.

73

- Within our investigation of the role of individual differences as the second aspect and contribution of this thesis, we fail to observe any notable effect of cognitive style to accuracy, efficiency or even approach taken, which indicates that the construct of cognitive style, as the concern of the research question [RQ2.1], might not be useful for studying the phenomena at hand as a trait. However, it appears to be promising as a characterization of cognitive strategy inspired by the characteristics of the task at hand. For the second part of the research question [RQ2.2], we find that mathematics anxiety has a mild negative correlation with performance irrespective of representation methods. Furthermore, our investigation of arithmetic ability effect on accuracy in either representation [RQ2.3] seem to show no notable interaction. And finally, the question regarding followed method [RQ2.4], following a certain method when solving the goal model problem shows an effect on the accuracy of the participants responses as mentioned earlier. This is a strong finding which we believe motivates further explanatory work.

- Thirdly, we support the feasibility of considering "intuitiveness" as one of the possible comparison constructs for comparing modeling notations, defined as the ability of novice users of the modeling notation to correctly understand how they can use it. We operationalized the construct of intuitiveness by measuring agreement between inferences that the language designers consider valid and inferences participants make, as well as the time it takes for the latter to take place. We also include the option of measuring perceived intuitiveness via self-reporting the confidence of participants of their inferences in the experiment. The presence of detectable effects seems to indicate that these constructs are promising instruments for talking about and evaluating model quality.

- Finally, our understanding of intuitiveness appears to be consistent with theories of graph comprehension and conceptual model visualization, including Pinker's theory of graph comprehension [55] and Moody's concept of semantic transparency [51] . This contribution shed light on the possibility on adapting this theory on conceptual modeling in general, and goal modeling contribution links in particular.

5.2   Future Work

Many different proposals in the representation of contribution link notation exist in the literature. While this work has investigated a specific pair of representation approaches, it is still pertinent to attempt similar experimental work on other representation and semantic approaches, for example a comparison of the URN [6] versus the Giorgini et al. [26] rules of propagation or the comparison of different quantitative approaches introduced in those two frameworks.

Future work can aim at the mechanics of the specific decision making task within goal models; particularly the distinction between developing a mental model, i.e., a theory of how the representation "works" versus executing the theory to perform useful inferences, which may also be obstructed by representational, complexity and other factors. For the task, instruments that enhance explanatory analysis need to be devised beyond our black-box technique. For the purpose, we believe that protocol analysis [21] may turn out to be of value. Applying the concepts of protocol analysis could help track the thought processes of problem solving. This way, through experiment designs in which monitoring subjects is within the research capability, protocol analysis methods such as verbal reports, reaction times, error rates can help understand the

elements in which users struggle with in trying to solve goal model problems.

Focusing on the results at hand on goal models, although they inform goal modeling practice (e.g. modelers can prefer numeric from symbolic contribution links whenever all else is equal), different domains, larger and different models are to be considered for further investigation using a similar approach. In that context it is important to keep in mind that the ultimate goal behind testing the intuitiveness of goal modeling contribution link is to find an effective method to communicate these models between analysts and stakeholders. Through the experiment presented in this work, the decision-making exercises reflect problems that can be relatable to the subjects of the experiments. These exercises, however, does not necessarily reflect real cases which goal models are purposed to reflect within the field of Information Systems. Future work may look into expanding the method used in this work to real decision problems regarding those of system design, or apply real case-studies of goal model approaches.

Through utilizing Pinker's theory of graph comprehension, it is proven to be applicable to graphs represented as conceptual models as cognitive processes play a role in the way they are perceived. A potential approach for future studies can look into adapting the theory or equivalent ones to offer a descriptive basis elements and varieties of conceptual models in addition to what has been covered in this work regarding contribution links.

Furthermore, since the stakeholders play a role in this equation, further studies can expand the methods of this research on experimental subjects of different domains including business, medicine, science in order to investigate the versatility of the research framework and the usability of the investigated goal model approaches.

Looking into different factors that may affect the intuitiveness of goal model contribution link notations, this work mainly focused on individual differences — in particu-

lar, cognitive styles, mathematical anxiety and arithmetic abilities, leaving other probable influencing individual differences factors such as linguistic capabilities, working memory, or physical disabilities, for example, subjects to study in future work concerning this particular research topic. In addition to individual differences, further studies on design factors regarding the comprehension of goal model contribution links such as the size and complexity of the models is required to expand the applicability of these approaches. The reason we look into these factors is to investigate the possibility to put these factors into consideration when developing or improving more representational methods.

Importantly, rather than just understanding a specialized task within a specific notation, our long-term aim is to develop an analytical and empirical perspective and took-set transferable to the study of more popular classes of notation, such as business process or diagrams [1] or entity relationship diagrams [14].

All in all, the quality of conceptual modeling languages is an important criterion to be used for guiding the design of such languages for the purpose of finding the optimal language to mediate between analysts and stakeholders. As more empirical studies, comparison and framework proposals are presented in the literature, better and more useful modeling practices will become in all relevant domains.

Bibliography

[1] Business Process Model And Notation. https://www.omg.org/spec/BPMN/2.0/.

[2] GRL - Goal-oriented Requirement Language. https://www.cs.toronto.edu/km/GRL/.

[3] JobTestPrep. https://www.jobtestprep.co.uk.

[4] ALLINSON, C. W., AND HAYES, J. The Cognitive Style Index. *Journal of Management Studies 33* (1996), 119–135.

[5] ALOTHMAN, N., ZHIAN, M., AND LIASKOS, S. User Perception of Numeric Contribution Semantics for Goal Models : an Exploratory Experiment. *Lecture Notes in Computer Science 10650* (2017).

[6] AMYOT, D., GHANAVATI, S., HORKOFF, J., MUSSBACHER, G., PEYTON, L., AND YU, E. Evaluating goal models within the goal-oriented requirement language. *International Journal of Intelligent Systems 25*, 8 (2010), 841–877.

[7] AMYOT, D., AND MUSSBACHER, G. User Requirements Notation: The First Ten Years, The Next Ten Years (Invited Paper). *JSW 6* (2011), 747–768.

[8] ASHCRAFT, M. H. Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science 11*, 5 (2002), 181–185.

[9] BIRKMEIER, D., KLÖCKNER, S., AND OVERHAGE, S. An Empirical Comparison of the Usability of BPMN and UML Activity Diagrams for Business Users. In *ECIS* (2010).

[10] BRESCIANI, P., PERINI, A., GIORGINI, P., GIUNCHIGLIA, F., AND MYLOPOULOS, J. Tropos: An Agent-Oriented Software Development Methodology. *Autonomous Agents and Multi-Agent Systems 8*, 3 (may 2004), 203–236.

[11] BROWN, V., CRONIN, M., AND BRYANT, D. Test of Mathematical Abilities – Third Edition (TOMA-3).

[12] CAIRE, P., GENON, N., HEYMANS, P., AND MOODY, D. L. Visual notation design 2.0: Towards user comprehensible requirements engineering notations. In *RE* (2013), IEEE Computer Society, pp. 115–124.

[13] CARVALLO, J., AND FRANCH, X. An empirical study on the use of i* by non-technical stakeholders: the case of strategic dependency diagrams. *Requirements Engineering* (2018).

[14] CHEN, P. P.-S. The Entity-relationship Model&Mdash;Toward a Unified View of Data. *ACM Trans. Database Syst. 1*, 1 (mar 1976), 9–36.

[15] CHUNG, L., A. NIXON, B., YU, E., AND MYLOPOULOS, J. *Non-Functional Requirements in Software Engineering*, vol. 5. 2000.

[16] COOLS, E., AND VAN DEN BROECK, H. Development and Validation of The Cognitive Style Indicator. *Journal of Psychology: Interdisciplinary and Applied 141*, 4 (2007), 359–387.

[17] CRONBACH, L. J., AND MEEHL, P. E. Construct validity in psychological tests. *Psychological bulletin 52*, 4 (1955), 281.

[18] CRUZ-LEMUS, J. A., GENERO, M., MANSO, M. E., MORASCA, S., AND PIATTINI, M. Assessing the Understandability of UML Statechart Diagrams with Composite states–A Family of Empirical Studies. *Empirical Softw. Engg. 14*, 6 (dec 2009), 685–719.

[19] DHILLON, M. K., AND DASGUPTA, S. Individual Differences and Conceptual Modeling Task Performance: Examining the Effects of Cognitive Style, Self-Efficacy, and Application Domain Knowledge. *Lecture Notes in Business Information Processing 81 LNBIP* (2011).

[20] EFKLIDES, A., PAPADAKI, M., PAPANTONIOU, G., AND KIOSSEOGLOU, G. Individual differences in school mathematics performance and feelings of difficulty: The effects of cognitive ability, affect, age, and gender. *European Journal of Psychology of Education 14*, 1 (1999), 57–69.

[21] ERICSSON, K. A. C H A P T E R 13 Protocol Analysis and Expert Thought : Concurrent Verbalizations of Thinking during Experts ' Performance on Representative Tasks.

[22] FIELD, A. *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. Sage Publications Ltd., 2013.

[23] FIGL, K., AND LAUE, R. Cognitive Complexity in Business Process Modeling. In *Advanced Information Systems Engineering* (Berlin, Heidelberg, 2011), H. Mouratidis and C. Rolland, Eds., Springer Berlin Heidelberg, pp. 452–466.

[24] FIGL, K., RECKER, J., AND MENDLING, J. A Study on the Effects of Routing Symbol Design on Process Model Comprehension. *Decis. Support Syst. 54*, 2 (jan 2013), 1104–1118.

[25] GIORGINI, P., MYLOPOULOS, J., NICCHIARELLI, E., AND SEBASTIANI, R. Reasoning with Goal Models. In *Proceedings of the 21st International Conference on Conceptual Modeling* (Berlin, Heidelberg, 2002), ER '02, Springer-Verlag, pp. 167–181.

[26] GIORGINI, P., MYLOPOULOS, J., NICCHIARELLI, E., AND SEBASTIANI, R. *Formal Reasoning Techniques for Goal Models*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 1–20.

[27] HADAR, I., REINHARTZ-BERGER, I., KUFLIK, T., PERINI, A., RICCA, F., AND SUSI, A. Comparing the comprehensibility of requirements models expressed in Use Case and Tropos: Results from a family of experiments. *Information and Software Technology 55*, 10 (2013), 1823–1843.

[28] HAMMOND, K. R., HAMM, R. M., GRASSIA, J., AND PEARSON, T. Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics 17*, 5 (1987), 753–770.

[29] HARRISON, A. W., AND RAINER JR, R. K. The influence of individual differences on skill in end-user computing. *Journal of Management Information Systems 9*, 1 (1992), 93–111.

[30] HAYES, J., AND ALLINSON, C. W. Cognitive Style and its Relevance for Management Practice. *British Journal of Management 5*, 1 (1994), 53–71.

[31] HMIELESKI, K. M., AND CORBETT, A. C. Proclivity for Improvisation as a Predictor of Entrepreneurial Intentions. 45–63.

[32] HOPKO, D. R., MAHADEVAN, R., BARE, R. L., AND HUNT, M. K. The Abbreviated Math Anxiety Scale (AMAS). *Assessment 10*, 2 (2003), 178–182.

[33] HORKOFF, J., AND YU, E. Interactive goal model analysis for early requirements engineering. *Requirements Engineering 21*, 1 (mar 2016), 29–61.

[34] HOUY, C., FETTKE, P., AND LOOS, P. Understanding Understandability of Conceptual Models – What Are We Actually Talking about? In *Conceptual Modeling* (Berlin, Heidelberg, 2012), P. Atzeni, D. Cheung, and S. Ram, Eds., Springer Berlin Heidelberg, pp. 64–77.

[35] HUMPHREYS, M. S., AND REVELLE, W. Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review 91*, 2 (1984), 153–184.

[36] JOŠT, G., HUBER, J., HERIČKO, M., AND POLANČIČ, G. An empirical investigation of intuitive understandability of process diagrams. *Computer Standards & Interfaces 48* (2016), 90–111.

[37] KOSSLYN, S. Understanding Charts and Graphs. *Applied Cognitive Psychology 3* (1989), 185–225.

[38] LARKIN, J. H., AND SIMON, H. A. Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science 11*, 1 (1987), 65–100.

[39] LIASKOS, S., DUNDJEROVIC, T., AND GABRIEL, G. Comparing alternative goal model visualizations for decision making: an exploratory experiment. In *SAC* (2018).

[40] LIASKOS, S., JALMAN, R., AND ARANDA, J. On Eliciting Preference and Influence Measures in Goal Models.

[41] LIASKOS, S., KHAN, S., SOUTCHANSKI, M., AND MYLOPOULOS, J. Modeling and Reasoning with Decision-Theoretic Goals. pp. 19–32.

[42] LIASKOS, S., MCILRAITH, S. A., SOHRABI, S., AND MYLOPOULOS, J. Representing and Reasoning About Preferences in Requirements Engineering. *Requir. Eng. 16*, 3 (sep 2011), 227–249.

[43] LIASKOS, S., RONSE, A., AND ZHIAN, M. Assessing the Intuitiveness of Qualitative Contribution Relationships in Goal Models: An Exploratory Experiment. In *Proceedings of the 11th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (Piscataway, NJ, USA, 2017), ESEM '17, IEEE Press, pp. 466–471.

[44] LIASKOS, S., AND TAMBOSI, W. Factors affecting comprehension of contribution links in goal models: An experiment. In *International Conference on Conceptual Modeling* (2019), Springer, pp. 525–539.

[45] LIU, L., AND YU, E. From Requirements to Architectural Design - Using Goals and Scenarios. *1st International Workshop From Software Requirements to Architectures, STRAW 2001* (2001), 1–9.

[46] LUCIA, A. D., GRAVINO, C., OLIVETO, R., AND TORTORA, G. Data Model Comprehension: An Empirical Comparison of ER and UML Class Diagrams. In *2008 16th IEEE International Conference on Program Comprehension* (jun 2008), pp. 93–102.

[47] MAIDEN, N., PAVAN, P., GIZIKIS, A., CLAUSE, O., KIM, H., AND ZHU, X. Making decisions with requirements: Integrating i* goal modelling and the AHP. Tech. rep., Essen, Germany, 2002.

[48] MARTIN, L. P. The Cognitive-Style inventory. *The 1989 Annual: Developing Human Resources 8*, 1962 (1989), 123–139.

[49] MENDLING, J., AND STREMBECK, M. Influence Factors of Understanding Business Process Models. In *Business Information Systems* (Berlin, Heidelberg, 2008), W. Abramowicz and D. Fensel, Eds., Springer Berlin Heidelberg, pp. 142–153.

[50] MESSICK, S. The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist 19*, 2 (1984), 59–74.

[51] MOODY, D. L. The " Physics " of Notations : Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. 756–779.

[52] MOODY, D. L., HEYMANS, P., AND MATULEVICIUS, R. Visual syntax does matter: Improving the cognitive effectiveness of the i* visual notation. *Requirements engineering 15*, 2 (2010), 141–175.

[53] MYLOPOULOS, J., CHUNG, L., LIAO, S., WANG, H., AND YU, E. Exploring Alternatives During Requirements Analysis. *IEEE Softw. 18*, 1 (jan 2001), 92–96.

[54] NIEL, J. Ten Usability Heuristics.

[55] PINKER, S., AND FEEDLE, R. *A theory of graph comprehension.* 1990, pp. 73–126.

[56] PURCHASE, H. C., WELLAND, R., MCGILL, M., AND COLPOYS, L. Comprehension of diagram syntax: An empirical study of entity relationship notations. *International Journal of Human Computer Studies 61*, 2 SPEC. ISS. (2004), 187–203.

[57] RECKER, J. C., AND DREILING, A. Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In *18th Australasian Conference on Information Systems* (Toowoomba, Australia, 2007), M. Toleman, A. Cater-Steel, and D. Roberts, Eds., University of Southern Queensland, pp. 356–366.

[58] REIJERS, H. A., AND MENDLING, J. A Study Into the Factors That Influence the Understandability of Business Process Models. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 41*, 3 (may 2011), 449–462.

[59] REIS, H. T., AND JUDD, C. M. *Handbook of research methods in social and personality psychology*. Cambridge University Press, 2000.

[60] RICHARDSON, F. C., AND SUINN, R. M. The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology 19*, 6 (1972), 551–554.

[61] ROSNOW, R. L., AND ROSENTHAL, R. *Beginning Behavioral Research: A Conceptual Primer*, 6 ed. Prentice-Hall, Inc, 2008.

[62] SCHALLES, C., CREAGH, J., AND REBSTOCK, M. Usability of Modelling Languages for Model Interpretation: An Empirical Research Report, 2011.

[63] SEIN, M. K., AND BOSTROM, R. P. Individual Differences and Conceptual Models in Training Novice Users. *Human-Computer Interaction 4*, 3 (1989), 171–195.

[64] SERRANO, M., CALERO, C., TRUJILLO, J., LUJÁN-MORA, S., AND PIATTINI, M. Empirical validation of metrics for conceptual models of data warehouses. pp. 506–520.

[65] SHANKS, G. Conceptual Data Modelling: an empirical study of expert and novice data modellers. *Australasian Journal of Information Systems; Vol 4, No 2 (1997) 4* (1997).

[66] SHARP, H., ROGERS, Y., AND PREECE, J. *Interaction Design: Beyond Human Computer Interaction*. John Wiley & Sons, Inc., USA, 2007.

[67] SHOVAL, P., AND FRUMERMANN, I. OO and EER Conceptual Schemas: A Comparison of User Comprehension. *Journal of Database Management - JDM 5* (1994).

[68] SUINN, R. Mathematics Anxiety Rating Scale. 0–1.

[69] SUINN, R. M., AND WINSTON, E. H. The Mathematics Anxiety Rating Scale. *Psychological Reports*, 92 (2003), 167–173.

[70] TABACHNICK, B., AND FIDELL, L. *Using Multivariate Statistics*, 6 ed. Pearson, 2012.

[71] VANDERFEESTEN, I., REIJERS, H. A., MENDLING, J., AALST, W. M., AND CARDOSO, J. On a Quest for Good Process Models: The Cross-Connectivity Metric. In *Proceedings of the 20th International Conference on Advanced Information Systems Engineering* (Berlin, Heidelberg, 2008), CAiSE '08, Springer-Verlag, pp. 480–494.

[72] WEINER, E. A., AND STEWART, B. J. *Assessing Individuals: Psychological and Educational Tests and Measurements*. Little, Brown, 1984.

[73] XUE, Y. Impacts of Cognitive Style and Strategic Reference Point on Decision-making Choices of Managers. *2012 International Conference on Management Science & Engineering* (2012), 7.

[74] YU, E. K.-H. Towards Modeling and Reasoning Support for Early-Phase Requirements Engineering. In *RE* (1997).

Appendix A

Survey Appendix

A.1 Consent Form

About this survey

Thank you for agreeing to participate in this study. It will take about 45 mins of your time and involves your observing a number of diagramatic models and answering questions about them. The data you will provide will help us understand better how a specific modeling language (goal models) can become more effective. Please keep reading for more information about the study.

Information about this study

Confirm you want to do this survey

Study Name: Comparing Representations and Semantics of Contribution Labels in Goal Models

Purpose of the Research: In Information Systems analysis, graphical goal models

constitute a popular way for acquiring and representing user requirements. Such models allow analysts to represent how goals of users are reduced into sets of alternative system functions that fulfill top level goals. Contribution links show how functional alternatives affect higher level goals of the stakeholders, allowing them thereby to make decisions. However, many ways for representing and assigning meaning to contributions have been proposed. Which one is better? In this research we aim at understanding which one is better in terms of naturalness, i.e. in terms of what untrained users expect the result of contribution combination to be based on how it presents itself in the diagram. We also aim at understanding if individual differences can act as predictors of how users understand and use the models.

What You Will Be Asked to Do in the Research: In this research you will be asked to perform the following tasks:

- Provide us some basic demographics: sex, age and educational background.

- Perform a pre-test that assesses your cognitive style.

- Watch one or more instructional videos on the task you will perform next.

- View a number of visual representations and answer comprehension or other questions pertaining to the visualizations.

- Describe in your own words how your worked to answer the previous questions.

- (if applicable and you consent) Have a concluding discussion with the PI or his Research Assistant to explain in your own words your opinion about the visualizations and how you worked with them.

Your participation will last between 30 and 45 mins for this pre-test and another 30-45 mins for the main test.

Risks and Discomforts: We do not foresee any risks or discomfort from your participation in the research.

Benefits of the Research and Benefits to You: Include a statement regarding any benefits of the research as well as benefits to the research participants.

Voluntary Participation and Withdrawal: Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer, to stop participating, or to refuse to answer particular questions will not influence the nature of the ongoing relationship you may have with the researchers and York University either now, or in the future. If you stop participating, you will still be eligible to receive the promised pay/compensation for agreeing to be in the project. In the event you withdraw from the study, all associated data collected will be immediately destroyed wherever possible.

Confidentiality: All personally identifying information you supply during the research – namely: name, address and audio recording – will be held in confidence and your name will not appear in any report or publication of the research. Your data will be collected though the use of this on-line instrument and, in the case of the follow-up interview, an audio recording device. This personally identifying data will be safely stored in a locked facility and a password protected digital medium (USB key or DVD) and only research staff/research team members will have access to this information. The personally identifying data will be destroyed by the end of 2020. Confidentiality will be provided to the fullest extent possible by law.

Your non-personally identifying responses – including: demographics (sex, age, educational background), responses to tests and exercises as well as (anonymized, wherever necessary) transcriptions from the audio recordings – may be published on the PI's web-sites (personal, lab) and shared with the research community as data files.

The PI will screen the data and use his own judgement to identify accidentally or indirectly personally identifying features and will redact the parts in question or, if necessary, the entire response prior to availing it to third parties. The anonymized non-personally identifying data collected in this research project may also be used by members of the research team in subsequent research investigations exploring similar lines of inquiry.

The researcher(s) acknowledge that the host of the online survey may automatically collect participant data without their knowledge (i.e., IP addresses.) Although this information may be provided or made accessible to the researchers, it will not be used or saved without participant's consent on the researchers' system. Further, because this project employs e-based collection techniques, data may be subject to access by third parties as a result of various security legislation now in place in many countries and thus the confidentiality and privacy of data cannot be guaranteed during web-based transmission.

Legal Rights and Signatures: I consent to participate in the study "Comparing Representations and Semantics of Contribution Labels in Goal Models" conducted by Sotirios Liaskos. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by consenting to participate.

☐ I understand the conditions of this study

Important data protection information When you start, this survey will store your answers, your internet address, and browser information on the PsyToolkit server. The responsibility for this survey rests entirely with the researcher(s) listed above. Click here if you do not want to participate now.

A.2 Pre-test

A.2.1 Mathematical Anxiety

Mathematical Anxiety is measured by the following items provided by Hopko et al. (2003):

Item	Low Anxiety	Some Anxiety	Moderate Anxiety	Quite a bit of Anxiety	High Anxiety
Having to use the tables in the back of a mathematics book.					
Thinking about an upcoming mathematics test one day before.					
Watching a teacher work an algebraic equation on the blackboard.					
Taking an examination in a mathematics course.					
Being given a homework assignment of many difficult problems which is due the next class meeting.					
Listening to a lecture in mathematics class.					
Listening to another student explain a mathematics formula.					
Being given a 'pop' quiz in a mathematics class.					
Starting a new chapter in a mathematics book.					

A.2.2 Mathematical Exercises

Mathematical Ability is measured by a custom-made exercises:

Can you now perform a set of arithmetic exercises? In the following screens you will be presented with simple mathematical problems. Can you respond with the correct solution? Can you do this as quick as you possible?

ADDITION	
0.76 + 0.19 =	....
0.47 + 0.73 =	....
0.77 + 0.28 =	....
0.89 + 0.66 =	....

Exercises will start getting more difficult involving MULITPLICATION and DIVISION. It is important that you provide an answer quickly (i.e., within 30 seconds). Approximate when you can't compute! Remember: no pen, paper, calculators or other aids are allowed.

MULTIPLICATION	
0.28 x 0.27 =	....
0.95 x 0.49 =	....
0.15 x 0.34 =	....
0.74 x 0.45 =	....
DIVISION	
0.46 / 0.62 =	....
0.44 / 0.54 =	....
0.37 / 0.88 =	....
0.18 / 0.73 =	....

For the following exercises you are asked to choose the larger of two quantities. It is important that you provide an answer quickly (i.e., within 30 seconds). Please remember that aids are not allowed.

WHICH ONE IS LARGER?	
ADDITION	
0.92 + 0.16	0.25 + 0.89
0.36 + 0.72	0.59 + 0.54
0.55 + 0.36	0.72 + 0.24
0.85 + 0.51	0.37 + 0.95
SUBSTRACTION	
0.66 - 0.46	0.81 - 0.66
0.48 - 0.11	0.61 - 0.29
0.75 - 0.35	0.42 - 0.06
0.81 - 0.12	0.93 - 0.28
MULTIPLICATION	
0.51 x 0.41	0.32 x 0.82
0.48 x 0.29	0.12 x 0.79
0.03 x 0.71	0.37 x 0.19
0.19 x 0.83	0.32 x 0.36
DIVISION	
0.54 / 0.73	0.69 / 0.88
0.25 / 0.98	0.11 / 0.55
0.52 / 0.72	0.59 / 0.87
0.54 / 0.73	0.69 / 0.88

A.3 Goal Model Survey

Knowledge of Goal Models

Experience of Goal Model was measured through the following question:

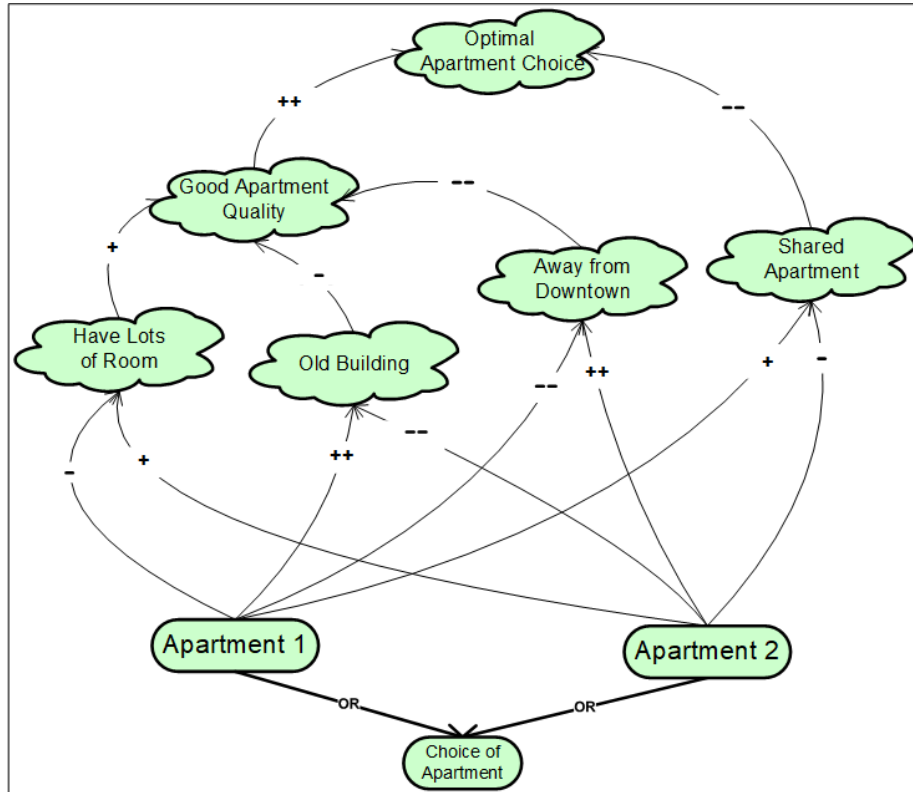| What is your experience with goal models?	
	I have never heard of them.
	I might have heard of them but have never seen any.
	I have viewed a goal model before.
	I have created a goal model.

Instructional Videos

- Decision Making Video Link: `https://youtu.be/7t6ElFksI4U`

- Goal Models Video Link: `https://youtu.be/qSdU8r_wBAA`

A.3.1 Goal Model Exercises

In each of the following screens you will be asked to solve a decision problem, presented to you as a goal model. Each model will have two or more alternatives. However, based the information in the goal model, one alternative is better than the others. Can you find which one? Can you find it quickly?

Apartment Choice

According to the above model, which apartment is the optimal choice?

- Apartment 1

- Apartment 2

Usage of Method

Method is measured through the following question:

What is your experience with goal models?	
	No, I used my intuition.
	Yes, I used a specific method.