NORTHWESTERN UNIVERSITY

A Cognitive Model of Recognition-Based Moral Decision Making

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Morteza Dehghani

EVANSTON, ILLINOIS

DECEMBER 2009

ABSTRACT

A Cognitive Model of Recognition-Based Moral Decision Making

Morteza Dehghani

The study of decision making has been dominated by economic perspectives, which model people as rational agents who carefully weigh costs and benefits and try to maximize the utility of every choice, without consideration of issues such as cultural norms, religious beliefs and moral rules. However, psychological findings indicate that in many situations people are not rational decision makers as defined by the economic theories. One of the domains in which traditional cost-benefit models fail to predict human behavior is the domain of moral reasoning.

This work presents the first computational model of recognition-based moral decision making, MoralDM, which integrates several AI techniques in order to model recent psychological findings on moral decision making. MoralDM uses a natural language system to produce formal representations from psychological stimuli, reducing tailorability. The impacts of secular versus sacred values are modeled via qualitative reasoning, using an order of magnitude representation. MoralDM uses a combination of first-principles reasoning and analogical reasoning to model the recognition-based mode of decision making.

The results of MoralDM experiments provided the impetus to further examine the role of cultural narratives and analogical reasoning on moral decision making. This work examines whether the processes by which core cultural narratives are applied in people's lives follow the principles of analogical retrieval and mapping. In particular, it examines how analogical accessibility and alignability influence the use of canonical moral narratives. The results of a

series of experiments performed among Iranian and American participants are reported and these results are simulated using MoralDM.

The last contribution of this thesis is regarding the use of structured qualitative representations and analogical generalization in modeling the similarities and differences in causal reasoning for biological kinds. The individual models built from transcript data are used to construct generalizations, which are tested both by inspection and by creating a classifier to distinguish models based on the culture and the level of expertise of the participants.

Overall, this thesis argues for the importance of highly structural representations in conjunction with analogical reasoning for capturing and modeling some effects of culture on cognition.

Acknowledgments

I would like to thank my advisor Ken Forbus for guiding me throughout this journey. Ken never lost hope in me and always showed interest in my ideas. He taught me to enjoy research and to always focus on the central research problems of my work. Without Ken's inspiration and support I would have never finished this dissertation.

I am grateful for my thesis committee. I take pride in having arguably three of the 10 most influential cognitive scientists on my committee. Doug Medin is the platonic form of an academic. Doug's commitment to science solely for the sake of science is truly inspiring. Dedre's encouragements and inspirations are limitless. You walk with an idea into her office, and two hours later, you walk out feeling as if you have discovered the most ground-breaking theory in cognitive science. Throughout the past four years, Tom Hinrichs' insightful questions engaged me to seek out some of the fundamental theoretical issues pertinent for my interdisciplinary research. Lastly, Bryan Pardo provided the moral support and encouragement that I needed to believe that there is some light at the end of the tunnel… some!

I would like to thank past and current members of the Qualitative Reasoning Group. Particularly, Praveen Paritosh and Andrew Lovett, for constantly questioning my ideas and strongly challenging my knowledge about AI and cognitive science. The help and constant support of Matthew Klenk, Emmett Tomai and Kate Lockwood gave me the drive necessary to reach the finish line. Working with this group of highly intelligent and encouraging individuals made the experience of graduate school bearable and more enjoyable. Also, I would like to thank Francisco Iacobelli for his statistically significant help with analyzing my data. Finally, working with Rumen Iliev and Sonya Sachdeva inspired me to pursue research in the field of cognitive psychology.

I would like to thank Hamed Ekhtiari and members of his lab for their help in my data collection in Iran. I am grateful to Janet Afary and Stefan Kaufmann for all their support throughout the past years. Also, I would like to thank my long time teacher and educator Mr. Rezaieyan for inspiring me to achieve higher education.

I am grateful to my friends all over the world. Please know that your friendship and your support mean a lot to me. I would especially like to thank Pouya Abghari, Ala Alavi, Shirin Alaie, Foad Dabiri, Hamidreza Maleki, Kaveh Naeeni, Shakila Shayan and Damon Soraya.

My family in Chicago, Azar Khounani and Hamid Akbari, truly helped me in the moments that I really missed home and needed a sense of family. Since I was a child, Hamid set an example for the person that I aimed to become. His dedication to politics, education and activism is truly inspiring.

I am grateful to my grandparents, Ali Akbar Khounani, Sorahi Nezam Hashemi, Morteza Dehghani and Zahra Dehbashi. Baba Ali and Maman Soory lovingly gave me shelter and took care of me the best that they could in the first years that I moved to the US. I never got to know Madar as much as I loved to and Baba Seyyed passed away right before I was born. However, I am proud of them and love them for who they were and what they represented.

Shirin Birjandi provided much needed encouragement, support and love in the last two years. She completely changed the experience of graduate school for me, making what was unbearable, enjoyable. I am most grateful to her for reminding me of the joys of the little things in life, things I often tend to forget. I would like to thank her for understanding me and standing by me.

I would like to thank my sister Kianoush Dehghani for teaching me how to dream, for boldly encouraging my intellectual pursuits, for the constant questioning of my beliefs, for teaching me to be different and for being the person that she is. I would like to thank her for being in my life, so that I can experience what it means to be a dreamer and what it means to experience dreaming of a better world and what it takes to strive for that dream.

Most importantly I would like to thank my loving parents, Akram Khounani and Hossein Dehghani. They taught me how to fight, seek higher and not be satisfied with any status quo. They sacrificed all through their lives to provide me what was not provided for them. Not this work, nor any other future achievement in my life, can repay what they went through for me to get to the position that I am. What I have achieved, what I shall achieve, is because of them and is will always dedicated to them.

This dissertation is dedicated to

Mohsen Ruholamini (1984-2009),

and all the brave souls of Iran's Green Movement.

Your stories of heroism and sacrifice will live for centuries,

inspiring each generation's struggle for freedom.

Table of Contents

List of Figures

List of Tables

1. Introduction

Decision making refers to the cognitive task of evaluating and selecting from a set of options in order to satisfy a set of goals. The study of decision making has been dominated by economic perspectives, which model people as rational agents who carefully weigh costs and benefits and try to maximize the utility of every choice, regardless of the domain and the context of that decision. These theories, which focus on the utility of outcomes, assume that people make consistent optimal choices across settings in order to maximize their utility (von Neumann & Morgenstern, 1944; Edwards, 1954). In other words, the rational actor theories essentially reduce the process of decision making to maximization of utility functions, without consideration of issues such as cultural norms, religious beliefs and moral rules which exist outside the market (Iliev, Sachdeva, Bartels, Joseph, Suzuki, & Medin, 2009). However, psychological findings indicate that in many situations people are not optimal nor rational decision makers as defined by economic theories. One of the domains in which the rational actor perspective fails to explain human behavior is that of moral decision making. A body of research illustrates that in the presence of certain moral values, people tend to focus on the obligations and duties outlined in their culture regarding that moral value and consequently are less concerned about the outcome utility of their choice. Other psychological research has shed light on the different processes involved in human decision making. Among other factors, these studies emphasize the importance of analogical and causal reasoning in conjunction with the strong influence of culture on reasoning and decision making.

The focus of this thesis is on modeling the role of such factors on moral decision making. Specifically, I describe the first computational model of recognition-based moral decision making which integrates analogical, qualitative and first principle reasoning techniques from AI and cognitive science. Moreover, I investigate the role of cultural narratives on moral decision making and model these effects. The last part of this thesis examines the role of culture on causal reasoning about biological kinds using cognitively plausible computational models of analogy and categorization. I argue for the importance of highly structural representations, analogical and qualitative reasoning in computational modeling of cultural phenomena. Before discussing the claims and contribution of my thesis, I review the existing models of decision making in AI and discuss their advantages and their shortcomings.

1.1. AI Models of Decision Making

Current AI models of decision making fail to capture many psychological properties of human decision making. These models focus on utility maximization, and use expensive reasoning methods or impoverished domain representations to model some aspects of decision making. Moreover, none of these models exploit cultural norms and stories or known heuristics that people use when making decisions. I argue in this thesis that moving away from utility based models and applying integrated techniques help us both understand the underlying processes of human decision making, and give our decision making models more power, to tackle a broader range of problems.

Computational models of decision making in AI can be categorized into three major types depending on the underlying formalism used in the models. The next section discusses these classes of models, their shortcomings, and advantages.

1.1.1. Connectionist Models

The main idea behind connectionist approaches to decision making is that decisions are based on "the accumulation of the affective evaluation produced by each action" (Busemeyer and Johnson 2004) until a threshold is satisfied. In these approaches, the action activation that first reaches the threshold is chosen by the model. Decision Field Theory (Busemeyer and Johnson 2004) uses a sequential sampling process to make decisions. In this model, the "attention node" is connected to different "action nodes" and over time different affective values are calculated for the attention node, which in turn produces valences. These valences determine the preference for an action. An action whose preference reaches a given threshold is then chosen. The main difference between Usher and McClelland's (2001) Competing Accumulator Model and Decision Field Theory is that the former adopts Tversky and Kahneman's (1991) loss aversion hypothesis, which states that disadvantages have larger impacts than advantages. ECHO (Guo and Holyoak 2002), adapted from Thagard and Milligram (1995) connectionist model, is based on similar hypothesis as these other models. The main difference between ECHO and other models is that it contains a special node called the "external driver" which models the goal of making a decision. When a decision is presented, the driver node gets activated and in turn activates the attribute nodes, and the process of assigning valences is then repeated.

1.1.2. Probabilistic Models

These approaches mainly focus on decision making under uncertainty and perform probabilistic inference given a set of evidence. The majority of these models are not built upon psychological findings on decision making. MINERVA-DM: A Memory Processes Model for Judgments of Likelihood (Dougherty, Gettys and Ogden 1999) is a probabilistic model which accounts for some psychological findings on judgment phenomena, including frequency judgments, conditional likelihood judgments and base-rate neglect. This model is based on the MINERVA 2 memory model (Hintzman, 1984) which has been used to study various phenomena in memory retrieval.

Recently there has been a lot of work on Bayesian decision making in the medical domain (Ashby and Smith 2000, Parmigiani 2002, Sox, et al. 2007). These methods focus on evidence-based medicine and the uncertainty associated with each piece of evidence. The proposed models are mainly used by medical doctors for diagnosis of different diseases given sets of partial evidence.

Also, a few computational models for juror decision making have been proposed. Kerr (1993) proposes a stochastic model of juror decision making which is used to infer some properties of jurors' decisions such as their decision criterion and their evaluation of the importance of each piece of evidence. The Story Model of juror decision making (Pennington and Hastie 1993) is an explanation based model which uses evidence learned through the trial and computes expectations about what facts are needed to make the case a complete story.

1.1.3. Qualitative Models

Qualitative models of decision making started receiving significant attention in the 1990's. Some of these models include: Pearl (1993), Wilson (1995), Bonet and Geffner (1996) and Dubois and Prade (1998). What most of these models have in common is a method for representing degrees of preferences, beliefs and goals in qualitative measurements. Both Pearl's and Wilson's methods use order of magnitude reasoning to calculate utility ranks. Bonet and Geffner use ordinal relations to represent the importance of goals. These models take as input a set of propositions defining the situation, also a set of goals and goal rankings. Then a set of actions and action rules are applied to the inputs, and output situations are calculated. Based on goal preferences, the plausibility of each choice is then computed and choices are ordered according to their level of plausibility and utility.

1.1.4. Shortcomings of Existing Computational Models of Decision Making

As discussed previously, recent psychological results have shed light on the process of human decision making by showing predictable violations of axioms of economic theory. The majority of AI models operate based solely on utility economics. In other words, the main criterion that these models use is finding the choice which maximizes the utility of the agent. Moreover, these methods violate and/or fail to capture important psychological findings on human decision making. For instance, the majority of these models operate on impoverished domain representations. Both connectionist models and propositional Bayesian networks lack the level of expressiveness needed to articulate many aspects of human reasoning. These models do not reflect on previous experiences, background knowledge or the culture of the decision maker. Many of the methods used in traditional AI models of decision making require expensive computational inference methods or very large search spaces, which result in the system becoming intractable when dealing with real world decision making problems.

1.2. Claims and Contributions

As discussed previously, the overarching theme of this thesis is the significance of highly structural representations, plus analogical and qualitative reasoning, for modeling the effects of culture on cognition. Specifically, the first part of this thesis focuses on the utility of integrating these factors in modeling recognition-based moral decision making. In the second part, I argue for the important role of analogical reasoning in moral decision making. Lastly, I demonstrate the effectiveness of analogical reasoning along with structural qualitative representation for capturing the differences in causal ecological reasoning between two cultures and also between the expertise levels of the participants.

The first contribution of this thesis is a cognitively motivated intergraded model of recognition-based moral decision making called *MoralDM*. Psychological findings illustrate that when making decisions, we employ several different processes and strategies for solving the task at hand (Payne, Bettman, & Johnson, 1992), and the preference for the use of these strategies is a function of experience (Fong, Krantz and Nisbett 1986, Kruglanski 1989, Larrick, Morgan and Nisbett 1990). Although the effects of many of these processes on decision making has been examined and modeled, the utility of integrating these processes has not been studied in detail. Integrated reasoning approaches can help solve some of the shortcomings of existing models of decision making in AI. In this thesis I argue that moving away from utility based models, and applying integrated techniques, each having complementary strengths and weaknesses, help us both study the underlying processes of human decision making, and give our decision making models the ability to tackle a broader range of problems.

MoralDM intergrades several AI reasoning methods to model psychological findings about utilitarian and deontological types of reasoning. Deontological reasoning refers to the type of reasoning which focuses on acts rather than outcomes of decisions. It uses a natural language system to produce formal representations from psychological stimuli, reducing tailorability. It employs both first-principles reasoning and analogical reasoning to model known findings on moral decision making and compare previously solved cases and cultural stories to novel situations. The ability to use known solved cases and cultural stories in the decision making process helps the system model the recognition-based mode of decision making. The different impacts of secular versus sacred values (those which focus on obligations and moral rules) are modeled via qualitative reasoning, using an order of magnitude representation. I evaluate MoralDM on stimuli taken from a series of psychology experiments.

It is well known that analogy plays an important role in the process of decision making. However, this role has not yet been systematically examined in the domain of moral decision making. The second claim of this thesis is regarding the use of analogical reasoning in understanding novel moral situations. I examine whether the processes by which core cultural narratives are applied in people's lives follow the principles of analogical retrieval and mapping. In particular, I demonstrate how analogical accessibility and alignability influence the use of canonical moral narratives. I also show that access to different moral stories results in differences in moral preference across cultures. I report on the results of three experiments performed among

Iranian and American participants. My results indicate that analogical accessibility to cultural narratives that are similar in structure to a given dilemma is the differentiating factor in our participants' responses across the different variants and between the two cultural groups.

The last contribution of this thesis is regarding the use of structured qualitative representations and analogical generalization in modeling the similarities and differences in causal reasoning for biological kinds based on the culture and the level of expertise of the participants. The Qualitative Concept Map system is used for modeling and analyzing transcripts of interviews conducted with these groups. This system is an environment in which qualitative representations can be used to explore mental models, enabling different types of reasoning and simulations to be performed on these models. The individual models created from the transcript data are used to construct generalizations for the groups, which are tested both by inspection and by creating a classifier to distinguish models from these two cultures. My system successfully classified models according to cultural group membership. Furthermore, it was able to automatically categorize Menominee models based on the level of the expertise of the participants.

1.3. Organization

Chapter 2 reviews the background psychology research on decision making. Included in this discussion is an overview of moral decision making, sacred values and quantity insensitivity, followed by an overview of the recent research on modes of decision making. I also discuss the role of analogy in decision making and close by reviewing the relevant research on the influence of culture on reasoning and decision making.

Chapter 3 outlines the details of a cognitive model of recognition-based moral decision making called MoralDM. In this chapter I review the reasoning techniques used in each module of MoralDM and discuss a series of experiments performed on the system.

Chapter 4 focuses on the role of analogy and cultural narratives in moral decision making. In this chapter, using a series of cross-cultural psychological experiments, I examine whether the processes by which core cultural narratives are applied in people's lives follow the principles of analogical retrieval and mapping.

Chapter 5 discusses a cognitive scientist friendly modeling tool, called the Qualitative Concept Map system, which is used to model transcripts of interview data from a fieldwork study. It also examines the use of qualitative representations and analogical generalizations in modeling the similarities and differences in causal reasoning for biological kinds between two cultures.

Chapter 6 summarizes the claims of this thesis and discusses some future directions.

2. Background

This thesis is built upon findings on decision making and analogical processing in psychology and knowledge representation, logical, analogical and qualitative reasoning in artificial intelligence. In this chapter, I present the background research concerning decision making which is used throughout this dissertation. The rest of the background is discussed in the introductions of Chapters 3, 4 and 5. This chapter is organized as follows. First, I discuss the research on moral decision making. Next, I review how protected values result in quantity insensitivity, followed by a summary of psychological findings on modes of decision making. Next, I discuss the role of analogy in decision making. Finally, I close by briefly covering some relevant research on cultural differences in decision making.

2.1. Moral Decision Making

As discussed in the previous chapter, the field of decision sciences has been dominated by economic theories which model people as rational agents carefully weighing risks, costs and benefits of every option before making decisions. This rational actor model assumes that agents' main strategy in decision making is utility maximization, regardless of the context and the domain of the decision. However, psychological findings have shed light on the process of human decision making by showing violations of these axioms in different domains. Early findings by Kahneman, Slovic and Tversky (1982) indicate that when making decisions, people heavily rely on biases and heuristics which often result in systematic violations of axioms of rational decision making. Other researchers have also shown that many other factors such the similarity effect (Tversky 1972) and the attraction effect (Huber, Payne and Puto 1982) change the preference of decision makers in ways which cannot be explained by economic theories. A more recent body of psychological research argues that a single process or a single mode of decision making cannot capture the full spectrum of human decision making (Bennis, Medin, & Bartels, in press; Hastie, 2001). One of the domains in which traditional normative cost-benefit models fail to predict human behavior is the domain of moral reasoning (see Bennis, Medin and Bartels in press for a review). Psychological evidence indicate that people facing moral dilemmas often do not act in utilitarian ways, or in other words, they do not follow the dictates of cost-benefit analysis (CBA). The conflict between these normative outcomes and intuitive judgments suggested the existence of *protected* (Baron & Spranca, 1997) or *sacred values* (Tetlock, 2000), which are not allowed to be traded-off, regardless of the consequences. Baron and Spranca (1997) define protected values as "… those that resist trade-offs with other values, particularly with economic values" and argue that these protected values "arise out of deontological rules about actions rather than outcomes". A similar trade-off blockage was proposed by Tetlock (2003), who defines sacred values as "those values that a moral community treats as possessing transcendental significance that precludes comparisons, trade-offs, or indeed any mingling with secular values". These sacred/protected[1] values outweigh economic ones (Tetlock, 2003) as they "incorporate moral beliefs that drive action in ways dissociated from prospects for success" (Atran, Axelrod, & Davis, 2007).

Consider the starvation scenario (from Ritov and Baron 1999) below:

[1] These two term will be used interchangeably throughout this thesis

> A convoy of food trucks is on its way to a refugee camp during a famine in Africa. (Airplanes cannot be used.) You find that a second camp has even more refugees. If you tell the convoy to go to the second camp instead of the first, you will save 1000 people from death, but 100 people in the first camp will die as a result.
>
> Would you send the convoy to the second camp?

While the normative CBA decision would send the convoy to the second camp, since this transfer offers more overall utility, 63% of the participants chose to not transfer them. People who have sacred values tend to reject trade-offs and often show strong emotional reactions, such as anger and disgust, when these values are challenged (Tetlock, 2003). Some decision scientists have proposed that people with protected values may treat these values as goods with infinite utilities. As Baron and Spranca (1997) argue, if people did indeed assign infinite utilities to these values, they should focus all their efforts and resources trying to maximize these values, and neglect others. However, it clear that this is not the case and therefore the idea of infinite utility does not seem plausible.

2.1.1. Sacred Values and Quantity Insensitivity

Certain values, such as human life, are often considered to be protected since people strongly react to tradeoffs of these values. In the presence of sacred values, people tend to be less sensitive to outcome utilities in their decision making and more concerned about their moral obligations. In other words, sacred values concern acts and not outcomes. As a result, when dealing with a case involving protected values, people tend to be concerned with the nature of their action rather than the utility of the outcome. This results in decisions which are contrary to

CBA models. Baron and Spranca (1997) argue that when dealing with protected values people show *insensitivity* to quantity. That is, in trade-off situations involving protected values, they are less sensitive to the outcome utilities of the consequences. For example, in the above scenario people who considered life to be a sacred value could very well have perceived killing 100 lives as equally wrong as killing 1000 people, and as a result preferred the choice involving inaction (Iliev, Sachdeva, Bartels, Joseph, Suzuki, & Medin, 2009).

The amount of sensitivity (or insensitivity) towards outcomes vary with the context. Lim and Baron (1997) show that this effect varies across cultures. In addition to contextual factors, the causal structure of the scenario affects people's decision making. Waldmann and Dieterich (2007) show that people act more utilitarian, i.e., become more sensitive to the outcome utilities, if their action influences the patient of harm rather than the agent. They also suggest that people are less quantity sensitive when their action directly, rather than indirectly, causes harm. Bartels and Medin (2007) argue that the agent's sensitivity towards the outcome of a moral situation depends on the agent's focus of attention.

In the next chapter, I argue that varying degree of sensitivity towards outcome utilities can be accounted for using qualitative order of magnitude reasoning. I present a simplified version of Dague's (1993a) ROM(R) qualitative order of magnitude formalism which I use to capture these results.

2.2. Modes of Decision Making

As discussed previously, analytic strategies for cost-benefit based decision making have received the majority of attention within the decision sciences (Weber, Ames, & Blais, 2005; Weber, 1998). As a result, versions of CBA make up the dominant normative models of decision making. However, recent research suggests that decision makers employ a variety of different decision modes and this selection depends on the characteristics of the task and on the decision maker (Bennis, Medin, & Bartels, in press). Weber (1998, Weber et al. 2005) distinguishes the following decision modes:

1. Calculation-Based decision making: this mode involves analysis of outcomes of choice alternatives for determining the choice with the highest utility. Traditional CBA models have focused on this mode.
2. Recognition-Based decision making: the main focus on this mode is categorization and integration of previous experiences. In this mode, as soon as a situation is categorized as an analog of a previous experience, “if-then” rules are applied to pick the choice. This mode includes the following subtypes:
    a. Rule-based decisions: employing stored set of known rules and norms that dictate behavior.
    b. Case-based decisions: it is activated by analogical classification and is utilized when the present choice reminds the decision maker of past episodes.
    c. Role-based decisions: is focused on social roles and codes of conduct that dictate the option to be chosen.
3. Affect-Based decision making: this mode is focused on immediate, holistic and affective reactions to choice alternatives.

Bennis, Medin and Bartels (in press) propose the following additional modes, which are not mutually exclusive to the modes mentioned previously:

1. Imitation-based decision making

2. Advise-seeking
3. Identity-based decision making
4. Exploration based
5. Coherence-based decision making.

These modes realize different objectives and depending on the task can coexist to facilitate decision making. Weber et al. (2005) argue that the outcome of a decision is determined by the mode that is attended to the most and has received the greatest weight.

Bennis, Medin and Bartels (in press) argue that moral rules often conflict with the calculation-based mode of decision making since they are mainly concerned with duties and obligations rather than outcomes. However, these moral rules fit recognition-based, affect-based and other modes closely related to these two. This thesis focuses on recognition-based moral decision making and calculation-based decision making for situations not involving moral reasoning. A major component in recognition-based decision making is analogy. In the next section, I discuss the role of analogy in decision making.

2.3. Decision Making and Analogy

The link between analogy and decision making has been explored from various perspectives, including consumer behavior (Gregan-Paxton, 2001), political reasoning (May, 1973) and legal decision making (Holyoak & Simon, 1999). In the domain of political decision making, for example, the domino effect was broadly used as a frame to describe the establishing of new communist governments during the Cold War. Since the domino analogy implies that a single element could cause failure of the whole system, the US government decision makers would accept high costs to prevent this from happening even in countries of low strategic importance. Also, US policymakers considering intervention in Vietnam drew parallels with the Korean War. Because the Chinese joined the Korean War against the US, there was concern that US involvement in Vietnam would lead to a Chinese military response (Glad & Taber, 1990; Markman & Moreau, 2001).

When making a choice, a decision maker recognizes the current situation as analogous to some previous experience and draws inferences from her previous choices (Markman & Medin, 2002). Goldstein and Weber (1995) argue that the process of decision making is a constructive process in which the decision maker relies extensively on their background knowledge and previous experiences. Medin, Goldstone and Markman (1995) demonstrate that similarity-processing and decision making share important commonalities. These correspondences and parallels suggest common mental processes for the two tasks.

Kokinov (2005) demonstrates the use of analogy in risky decision making, showing how experiencing a single episode of success in risky decision making biases participants towards taking more risky decisions. Moreover, Petkov and Kokinov (2006) illustrate how structure mapping (Gentner, 1983) can account for some of contextual effects in decision making, such as the frequency effect (Parducci and Perret 1971), the concave form of the utility function (Kahneman and Tversky 1979) and the preference for the middle ratings (Petrov and Anderson 2005). In Chapter 4 I argue that the impact of cultural narratives on decision making can be captured using structure mapping theory.

2.4. Cultural Differences in Decision Making

Due to both the rapid globalization of commerce and the need to move from single population based models (Cole, 1996), the influence of culture on decision making has become a topic of interest for both psychologists and economists. Judgment and decision making researchers have highlighted a number of ways in which culture may influence decision making. The influence of culture on probability judgments, risk perception and risk preference has been extensively explored by researchers in different fields (see Weber and Hsee 2000 for a review). Probably the most well known results are the findings on cultural differences in judgments of risk (Weber & Hsee, 2000; Hsee & Weber, 1999; Weber & Hsee, 1998). Hsee and Weber (1997; 1998; 1999) have found that participants from collectivist cultures, such as East Asian cultures, are more risk-seeking because they have a larger social cushion to fall back on in case of loss. In addition, they have found that these differences are not due to differences in attitudes towards risk, rather it is something about how a risk is perceived and construed that differs. In addition, earlier findings on cultural differences in judgments of probability also show cultural differences. A number of studies conducted by Yates and colleagues (Yates, Zhu, Ronis, Wang, Shinotsuka, & Toda, 1989; Yates, Lee, & Shinotsuka, 1996; Yates, Lee, & Bush, 1997) have shown that people living in cultures that are organized collectivistically are more overconfident in their judgments than those in individualistically organized cultures. These striking differences on some of the most important issues in judgment and decision making indicate that the environmental context plays a large role in basic cognitive processes.

Cultural products created over generations and responsible for storing and transmitting cultural wisdom (Weber and Hsee, 1998), as well as the current economic or political status of a nation affect different aspects of people's judgment and decision making. These social and cultural causes are known to leave traces and are reflected by a variety of cultural products, such as the proverbs in that cultural (Weber and Hasee, 1998). In a recent paper, we (Ekhtiari, Behzadi, Dehghani, Jannati, & Mokri, in press) investigated the effect of frequency in contrast to amount of reward and/or punishment in risky decision making among Iranians and compared our results to the results of the same risky decision making task performed in Western countries. Our participants consistently chose more frequently and more rapidly from options which had less frequent but larger amounts of punishments in comparison to options which had more frequent punishments with smaller amounts. The main score of our participants was surprisingly lower than the results of studies done in the West (and Korea). Moreover, among the studies we surveyed significant differences only occurred when the two compared papers used participants of different cultures, with Americans always scoring higher than other cultures. We argue that some of these differences between our Iranian participants and their Western counterparts may be due to social-economic causes, such as multiple regime changes, years of war, instability in the social-political atmosphere and religious restrictions for gambling.

Most of the work that has been done looking at cross-cultural differences in morally-motivated decision making has been ethnographic in nature. Shweder et al. (1997) and Haidt, Koller and Dias (1993) have identified domains of moral decision making that are present in one cultural group but not in another. Domains such as respect for authority and the saliency of the distinction between purity and impurity are some that have been identified in helping people to characterize certain situations as morally tinged within one cultural group but not another.

3. MoralDM: A Computational Model of Recognition-Based Moral Decision Making

3.1. Introduction

While traditional models of decision making in AI have focused on cost-benefit analysis theories, there is considerable psychological evidence, as discussed in previous chapters, that these theories fail to capture the full spectrum of human decision making (Bennis, Medin, & Bartels, in press). In particular, cost-benefit analysis models fall short in predicting human behavior in the domain of moral reasoning. What makes this domain particularly interesting is that when faced with morally charged scenarios, participants often prefer actions which are "in ways dissociated from prospects for success" (Atran, Axelrod, & Davis, 2007). This preference is contradictory to economic theories of decision making which assume that people always make optimal choices in order to maximize their utility independently of the context of the scenario (von Neumann & Morgenstern, 1944; Edwards, 1954).

There is growing interest in various fields in understanding and modeling moral cognition. I believe computational modeling of moral decision making is important for two reasons. First, computational models of moral decision making can provide formal approaches for examining the underlying mechanisms behind moral cognition and provide formal vehicles for gaining new insight about how this task is performed in humans. Cognitive models of other aspects of our cognition, such as analogy, have helped us develop better understanding of those processes. In the same way, modeling moral reasoning can help us advance our understanding of moral cognition.

Second, as AI agents are becoming more sophisticated and more integrated in our daily lives, it is important that they can make distinctions between pure utilitarian and moral decisions. By modeling moral decision making, we can build agents that can better interact with, and aid, humans faced with an increasingly global community. Computational models which can account for culturally specific aspects of decision making could both help us better understand other cultures and also work as software agents for interacting with them. Moreover, with different types of agents and robots entering morally sensitive domains, such as the medical field and the military, it is important that they can make morally correct judgments.

In this chapter, I present the first cognitive model of recognition-based moral decision making, called MoralDM. Due to the complex nature of moral reasoning, an integrated approach is necessary for modeling this task. As I discussed in previous chapters, integrated approaches can provide the system with the ability to tackle a broader range of decision making problems. Furthermore, integrated approaches are more cognitively plausible, as research in cognitive psychology stresses the use of multiple processes in human decision making (Payne, Bettman, & Johnson, 1992). For capturing the relationship between utilities of choices, MoralDM uses qualitative representations, as they provide a useful commonsense approach for comparing values with varying degree of sensitivity.

MoralDM models psychological findings about utilitarian and deontological types of reasoning. The system takes as input a scenario in natural language and chooses a decision which is morally preferred for a given culture. This is achieved by integrating several AI techniques. To reduce tailorability, the system uses a natural language understanding system to assist in producing formal representations from the stimuli re-rendered in simplified English. MoralDM uses both first-principles and analogical reasoning to implement rules of moral decision making and utilize previously made decisions. The ability to use known solved cases and cultural stories in the decision making process helps the system model the recognition-based mode (Weber, 1998; Weber, Ames, & Blais, 2005) of decision making. The impacts of secular versus sacred values are modeled via qualitative reasoning, using an order of magnitude representation. Explanations for the decision made can be inspected by examining the reasoning trace of the system. I test this model on stimuli from a series of psychology experiments. Moreover, I use the model to make several theoretical claims. Also, I discuss how MoralDM can be used to capture and analyze cross-cultural differences in moral decision making.

This chapter is organized as follows. I begin by describing the overall architecture of the system followed by a detailed description of each of its modules. Next, I show that MoralDM can account for results from two psychological studies. Then, I discuss an analysis of its performance, demonstrating learning through accumulation of examples and the importance of integrated reasoning. Next, I argue how MoralDM in its current architecture can be used for modeling cross-cultural differences. Finally, I discuss related work.

3.2. MoralDM

Moral decision making is a complex reasoning process. In psychological studies, scenarios are presented to human participants in natural language. The research summarized in the previous chapter emphasizes the role of protected values on decision making and identifies a number of contextual factors which cause participants to become less sensitive to the outcome utilities of their decisions. Other research has highlighted the significant role of analogy in decision making (Markman & Medin, 2002). Consequently, a model of moral decision making needs to include natural language understanding, the facility to switch between types of reasoning based on context, a method for comparing outcome quantities that takes into account the effects of protected values, and the ability to utilize previous decisions or examples when reasoning about new situations.

MoralDM incorporates two mutually exclusive types of reasoning: utilitarian (cost-benefit analysis) and deontological. If there are no protected values involved in the case being analyzed and the case does not resemble any previously solved cases, MoralDM operates in the calculation-based mode and applies traditional rules of cost-benefit analysis (CBA) by choosing the action which provides the highest outcome utility. On the other hand, if MoralDM determines that there are sacred values involved or if finds a similar solved case or story in its memory, it operates in the recognition-based mode (case-based, rule-based or role-based). If the system determines that the case concerns a moral issue, deontological reasoning is applied, the system becomes less sensitive to the outcome utilities and it prefers inactions to actions, if actions would violate moral principles.

MoralDM has been implemented using the FIRE reasoning engine. The knowledge base contents are a 1.4 million fact subset of Cycorp's ResearchCyc[2] knowledge base, which provides formal representations of everyday objects, people, events and relationships. The KB also includes representations our group has developed to support qualitative and analogical reasoning. The KB provides a formal ontology that is useful for representing and reasoning about moral decision making scenarios. The rules included in each module have been implemented as

[2] research.cyc.com

backchaining rules in FIRE. I describe high-level rules that each module uses in its corresponding section. For more details about the rules and their implementation please refer to Appendix B.

Figure 1 provides an overview of the MoralDM architecture. To solve a given moral decision making scenario, MoralDM begins by using EA NLU, a natural language understanding system, to semi-automatically translate simplified English scenarios into predicate calculus. Given this representation, the presence of protected values and relevant contextual factors are computed via a fixed set of rules. A number of known protected values are stored in the KB. For a new scenario a set of rules are applied to decide whether the case includes sacred values or not. The orders of magnitude reasoning module (OMR) then calculates the relationship between the utility of each choice. Using the outcome of the orders of magnitude reasoning module, MoralDM utilizes a first-principles reasoning module (FPR) and an analogical reasoning module (AR) to arrive at a decision. The FPR module suggests decisions based on rules of CBA and moral reasoning. The AR module compares a given scenario with previously solved decision cases to suggest a course of action.

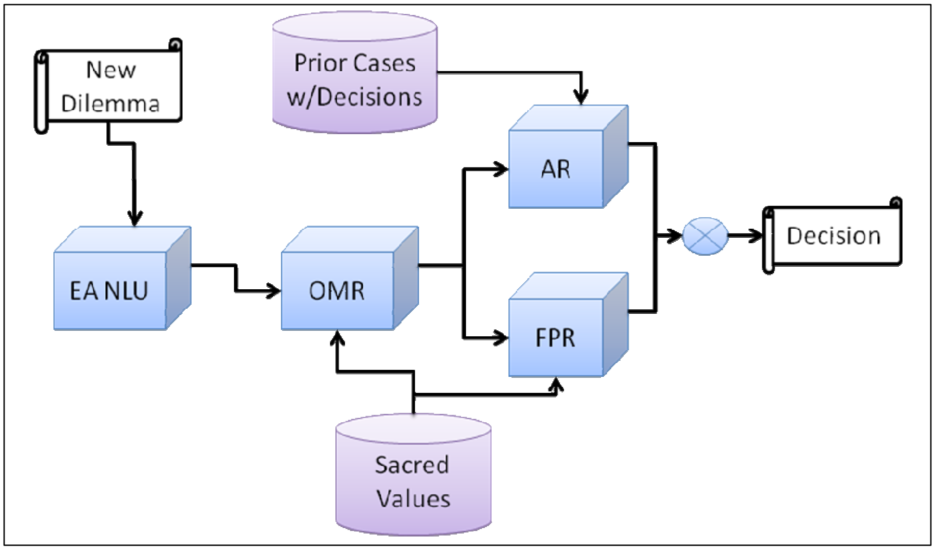


Figure 1: MoralDM architecture

The FPR and AR modules work in parallel and complement each other by providing support (or disagreement) for a decision. If both succeed and agree, the decision is presented. When one module fails to arrive at a decision, the answer from the other module is used. If the modules do not agree, the system checks the similarity score between the problem and its closest analog. If the two cases resemble each closely, then the system uses the derived answer from the analog. Otherwise, the system selects the FPR module’s choice. If both fail, the system is incapable of making a decision. After a decision is made for a given scenario, the decision along with reasons for choosing it and the mode in which it was made in are stored with the case itself in the case library for future use. The additional statements added to the case are either derived from the rules of the FPR module, or are information mapped from the base analog. Please refer to Appendix A for a sample solved case with the additional information saved with it. Storing previously solved cases in the case library enables the system to make decisions in more scenarios as it accumulates experience. Next, I discuss each module in detail.

3.2.1. Explanation Agent Natural Language Understanding system

The Explanation Agent Natural Language Understanding system (EA NLU) (Tomai, 2009a; Kuehne, 2004) component of MoralDM takes the input stimuli in natural language and constructs formal representations in predicate calculus. In typical cognitive modeling work, these representations are created by hand from the original texts, a process that is both labor-intensive and error prone. It also leads to the problem of *tailorability* – the possibility that representation choices were made to get a particular example to work, as opposed to being uniform, independently motivated conventions – since the simulation authors (or people working closely with them) do the encoding. EA NLU enforces a consistent knowledge framework and a set of automatic transformations, thus reducing tailorability and increasing the plausibility of the simulation results (Tomai & Forbus, 2009b). EA NLU has been used in several cognitive modeling experiments including MoralDM, conceptual change (Friedman & Forbus, 2008; Friedman, Taylor, & Forbus, 2009) and blame attribution (Tomai & Forbus, 2008).

EA NLU implements a *practical language understanding* approach to facilitate natural language input to cognitive simulations. This approach consists of three parts: 1) a large knowledge base with an expressive representation language, 2) simplified syntax and 3) a task-independent semantic interpretation process providing a query-driven interface (Tomai & Forbus, 2009b). For the knowledge base, EA NLU uses the contents of ResearchCyc plus our own extensions, as described above. This knowledge includes numerous *denotations* and *subcategorization frames* that link lexical terms to concepts in the Cyc ontology. These frames provide knowledge-rich semantics for words and common phrases.

In MoralDM experiments, inputs are dilemmas from the psychological literature, expressed in natural language. Unrestricted automatic natural language understanding is currently beyond the state of the art. Consequently, EA NLU uses a simplified syntax and operates semi-automatically, enabling experimenters to select among options presented by the system. This practical approach allows us to broadly handle syntactic and semantic ambiguities while constructing semantically expressive representations suitable for complex reasoning. EA NLU uses Allen's bottom-up chart parser (Allen, 1995) in combination with the COMLEX lexicon (Macleod, Grishman, & Meyers, 1998) and a simplified English grammar (Kuehne & Forbus, 2004). Each frame represents a case for the term encoded as predicate calculus with syntactic/semantic role variables. Roles are filled during the parsing process. Frames are filtered according to both case constraints and syntactic requirements explicitly included in the frame. Sentences within a stimulus are parsed separately. The resulting parse trees are presented, together with the semantic frames they entail, to the user in the interactive interface. The interface enables experimenters to quickly see the possible interpretations of their text, and select the desired alternatives. It provides real-time feedback about problems with unknown words and out-of-bounds syntax. Importantly, the experimenter is never asked to disambiguate the complex forms constructed by the composition, only per-term semantic frames and syntactic ambiguities such as prepositional attachments. Figure 2 contains a typical disambiguation choice as presented to the experimenter. This is a significant advantage over having experimenters construct representations by hand (Tomai, 2009a). The user can selectively include or exclude parse trees as well as individual frames. These selections serve as input to a transformation process using dynamic logic principles from Discourse Representation Theory (DRT) (Kamp & Reyle, 1993) to construct a description of the sentence content. The representation supports numerical and

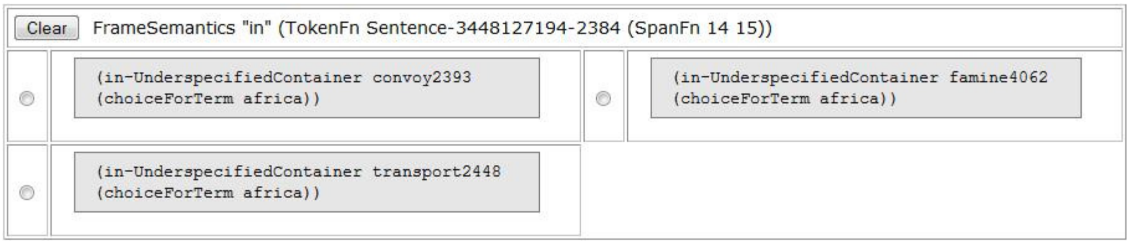


Figure 2: Disambiguation of semantic roles for the preposition "in"

logical quantification, negation, implication, modal embedding and explicit and implicit utterance sub-sentences. Explicit quantifiers, negation and implication are handled by constructing *discourse representation structures* (DRS). Modal operators are recognized for

```
A convoy of trucks is transporting food
to a refugee camp during a famine in
Africa. 1000 people in a second refugee
camp will die. You can save them by
ordering the convoy to go to that
refugee camp. The order will cause 100
people to die in the first refugee camp.
```

Figure 3: Starvation scenario in simplified

expected futures and possibilities (as presented from the speaker's point of view). The possible world of a modal operator is also represented by constructing a DRS. Utterances are represented using Cyc conventions for dialogue acts and information bearing microtheories.

Figure 3 contains the controlled language for the starvation scenario from a psychology study by Ritov and Baron (1999). Given these statements, EA NLU identifies events of transporting, famine, dying (1000 people), saving, ordering, going and dying (100 people) together with the two quantified sets of people, the convoy, food, two refugee camps and the proper name Africa. There is also an explicit reference to the listener, "you". Figure 4 contains the frame-based interpretation of the order.

```
(isa order131049 Ordering-CommunicationAct)
(performedBy order131049 you128898)
(recipientOfInfo order131049 convoy127246)

(infoTransferred order131049
 (and
  (isa refugee-camp129739 RefugeeCamp)
  (isa convoy127246 Convoy)
  (isa go129115 Movement-TranslationEvent)
  (primaryObjectMoving go129115 convoy127246)
  (toLocation go129115 refugee-camp129739)))
```

Figure 4: Filled semantic frame for ordering

This set of facts is contained within a DRS which is modally embedded with the operator `possible` in the root DRS for the scenario interpretation, indicating it is one outcome of the

```
(causes-SitProp order131049
 (and
  (isa set-of-people131188 Set-Mathematical)
  (cardinality set-of-people131188 100)

  (isa die131270 Dying)
  (forAll ?x
   (implies
    (elementOf ?x set-of-people131188)
    (and
     (isa ?x Person)
     (objectOfStateChange die131270 ?x))))
```

Figure 5: Predicate calculus for a quantified dying event

choice. In the starvation scenario, proper numerical quantification of the sets of people is an important part of understanding the choice presented by the scenario. Figure 5 contains the predicate calculus for 100 people dying, caused by the order given. Causal links are explicitly

```
(isa Sel131949 SelectingSomething)
(choices Sel131949 order131049)
(choices Sel131949 Inaction131950)
(causes-PropSit
  (chosenItem Sel131949 Inaction131950)
  die128829)
(causes-PropSit
  (chosenItem Sel131949 order131049)
  save128937)
```

Figure 6: Predicate calculus for the choice presented

stated between the order and the saving and the order and the second set of deaths. The abstraction of saving drives inferential attention to events in the description that the beneficiary may be being saved from. The expected future modality of the first set of deaths makes it a reasonable candidate. Based on the possible modality of the saving/ordering sequence, combined with the use of the explicit reference to the listener, the system infers an abstraction of choice being presented with known consequences resulting from both action and inaction. Figure 6 contains the inferred abstraction of choice and its causal consequences. Appendix A includes the full predicate calculus representation of two of the moral scenarios used in this chapter and an Iranian story used in later experiments.

3.2.2. Order of Magnitude Reasoning Module

As discussed in the previous chapter, in the presence of sacred values, people tend to be less sensitive to outcome utilities in their decision making (Baron & Spranca, 1997). In these circumstances, people are more concerned about the nature of their action, their duties and obligations. This insensitivity to outcome utilities sometimes results in decisions which are contrary to cost benefit analysis. People's degree of quantity sensitivity varies according to the causal structure of the scenario (Waldmann & Dieterich, 2007), participants' culture (Lim & Baron, 1997) and their focus of attention (Bartels & Medin, 2007). I claim that this variable degree of quantity sensitivity can be accounted for by using a qualitative order of magnitude representation.

We model quantity sensitivity by using Dague's (1993a; 1993b) qualitative Relative Order of Magnitude (ROM(R)) formalism. Order of magnitude reasoning is a form of commonsense reasoning. It provides the kind of stratification that seems necessary for modeling the impact of protected values on decision making. Raiman (1991) uses the analogy of a coarse balance to describe the intuitions behind order of magnitude reasoning: a course balance can weigh quantities with more or less precision. This precision level depends on the order of magnitude scale used to map quantities onto coarse values. He uses two granularity levels *Small* and *Rough* to build a multitude of order of magnitude scales. These two granularity levels provide three qualitative relations between quantities, which have been formally defined in FOG (Raiman, 1991). Both O(M) (Mavrovouniotis & Stephanopoulos, 1990; 1988; 1987) and ROM(K) (Dague, 1993b) are attempts to provide a more comprehensive order of magnitude formalism.

ROM(R) is the mapping of ROM(K) onto $\Re$ (real numbers). This order of magnitude formalism is the only system that guarantees validity in $\Re$. Given that in decision making scenarios utilities of choices are expressed in real numbers, ROM(R) seems to be the best order of magnitude formalism for our purposes. Some order of magnitude representations (e.g. FOG) do not allow values at different levels to ever be comparable. One of the features of ROM(R) is that it includes two degrees of freedom, k_1 and k_2, which for our purposes can be varied to capture differences in quantity sensitivity. Dague (1993b) defines four classes of relationship between two numbers: “close to”, “comparable to”, “negligible with respect to” and “distant from”. While FOG and O(M) fail to capture gradual change, the overlapping relations in ROM(K) allow a smooth, gradual transition between the states.

Although for engineering problems two degrees of freedom and four relations is quite useful, I believe for the task that I am interested in one degree of freedom and three binary relations are more plausible. Therefore, I implemented a simplified version of ROM(R) using one degree of freedom, k, resulting in three binary relations; almost equal, greater than, and orders of magnitude different. These three classes can be computed using the following rules:

- $A \approx_k B \Leftrightarrow |A\text{-}B| \leq k * \text{Max}(|A|,|B|)$
- $A <_k B \Leftrightarrow |A| \leq k * |B|$
- $A \neq_k B \Leftrightarrow |A\text{-}B| > k * \text{Max}(|A|,|B|)$

These relations respectively map to “close to”, “greater than” and “distant from”. k can take any value between 0 and 1. Figure 7 demonstrates the interval landmarks of the system. In order to provide smooth transitions between these qualitative intervals ROM(R) defines parameter ε. ε is the infinitesimal value which if added or subtracted from k would provide a smooth transition

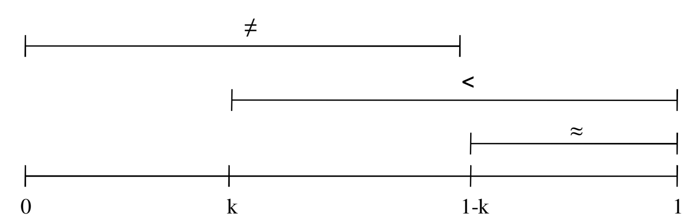


Figure 7: Interval landmarks

between the intervals. In our simplified version of ROM(R), when $k < ½$, ε is $k/(1 - k)$, and, when $k \geq ½$, ε is $(1 - k)/ k$. Quantity sensitivity can be varied by changing k: setting k to $k - \varepsilon$ shifts the relationship between the compared values and moves it from ≈ to < or from < to ≠ resulting in higher quantity sensitivity. On the other hand, setting k to $k + \varepsilon$ decreases the quantity sensitivity of the system as it shifts the relationships between values from < to ≈ or from ≠ to <. Depending on the protected values involved and the causal structure of the scenario, our system varies k to capture sensitivity towards the utility of the outcome.

The inputs to OMR include the protected values for the culture being modeled and the predicate calculus produced by EA NLU. The output is the order of magnitude relationship between outcome utilities. The highest level rule in the OMR module is `utilCalculation` which in turn calls `utilOfChoices`. The first rule calculates the expected utility of each choice by summing the utility of its consequences. This is done by first calling `choicesAre` (step 1) to determine the choices involved and `choicesAndConsequences` (step 2.1) to find the consequences of each choice. Next, the system determines whether the consequence of each choice is that of a promotion or of a prevention. Then, for each consequence of a choice, OMR uses its rules to ascertain if the outcome is positive or negative. This is done by calling `isNegativeUtil` (step 2.2). Next, OMR calls `cardinalityOfThingsEffectedByPromotion` and `cardinalityOfThingsEffectedByPrevention` to identify any sets whose cardinality matters in the decision (e.g., number of people at risk). Then, `findInitialK` (step 2.4) is called to calculate an initial value for k and ε.

After computing utilities, OMR adjusts the k value based upon the presences of protected value

Given: Protected values and output of EA NLU

1. Find the choices involved in the scenario
2. Calculate the utility of each choice:
   - 2.1. Determine the consequences of each choice
   - 2.2. Ascertain if a consequence has positive or negative utility
   - 2.3. For each consequence:
     - 2.3.1. For preventive outcomes of negative consequences and promotive outcomes of positive consequences: Calculate the cardinality of the things affected by the consequences
     - 2.3.2. For preventive outcomes of positive consequences and promotive outcomes of negative consequences: Calculate the cardinality of the things affected by the consequences and multiply it by -1
   - 2.4. Calculate k and e based on these values
3. Determine whether choices involve sacred values
   - 3.1. If so: set k to $k + e$
4. Determine the causal structure of the scenario:
   - 4.1. For direct causation or agent intervention: set k to $k - e$
   - 4.2. For indirect causation or patient intervention: set k to $k + e$
5. Based on the cardinality of sets and k, return the order of magnitude relationship between outcome utilities

Figure 8: Calculating the relationship between outcome utilities

and the causal structure of the scenario (e.g. agent v. patient intervention). Assuming that the relationship between the utilities, a and b, are "comparable", MoralDM sets k to $1 - (|a / b|)$. This results in the relationship between the utilities falling within <, right between ≠ and ≈ (Figure 7). If the decision involves a protected value for the modeled culture, setting k to $k + \varepsilon$ shifts the relationship between utilities from greater than to close to, resulting in the system being less sensitive to the numeric utility of the outcome. On the other hand, if the there are no protected values involved, the system substitutes k with $k - \varepsilon$ thereby making the system more quantity sensitive to the computed utilities.

OMR calls `involvesSacredValue` (step 3) to determine whether a particular choice involves protected values or not. In addition to protected values, the causal structure of the scenario affects k. OMR checks to see if the scenario contains patient intervention or agent intervention. It sets k to $k + \varepsilon$ in the first case, and to $k - \varepsilon$ in the second case, thereby making the system sensitive to the causal structure of the scenario, consistent with psychological findings (Waldmann & Dieterich, 2007). The system also checks for direct versus indirect causation. In the case of indirect causation, a higher degree of insensitivity is applied. OMR determines the causal structure of the scenario by calling rules such as `patientIntervention` and `doubleEffect` (step 4). Using the information about the causal structure of the scenario and whether or not protected values exist in the case, OMR determines the value of k by calling `determineK`. Next, `utilRelation` (step 5) is called which calculates the order of magnitude relationship between the outcome utilities. The result of this rule is made available to the FPR and AR modules. The rules for calculating utilities and determining causal structure are included in Appendix B.

Returning to the starvation scenario, there are two choices: ordering and inaction. For ordering, there are two consequences, 1000 people in the second camp will be saved and 100 people in the first camp will die. Consulting the KB, the system determines that dying has negative utility and saving positive, resulting in a choice utility of 900 for the ordering choice. Using the same procedure, the utility for inaction is calculated to be -900. Using the formula given above, k is initially set to 0 with $\varepsilon = 1$. Given that both choices involve agent intervention and indirect causation, there are no structural differences between the two choices. Therefore, the k value is set solely by the existence of protected values. In this case, causing someone to die is a sacred value resulting in k being set to $k + \varepsilon = 1$, therefore causing the system to act less quantity sensitive. Using our simplified version of ROM(R), the relationship between the utilities of the two choices is calculated to be ≈. On the other hand, if there had not been a protected value, the value of k would have remained 0 causing the relationship between the utilities to be ≠. The utilities, 900 and -900, and the computed relationship, ≈, are provided to FPR and AR.

3.2.3. First-Principles Reasoning Module

In many situations we apply rules to help us make decisions. “Rule following” is referred to situations when a series of rules (norms) are applied in order to guide the decision making process. It has been illustrated that decision makers often apply rules before engaging in other forms of reasoning (Mellers, Schwartz, & Cooke, 1998). Rule-based (and role-based) decision making are among the subtypes of recognition-based decision making (Weber, 1998; Weber, Ames, & Blais, 2005), where the decision maker has to first recognize the current situation as similar to a known norm or some previous case and then apply appropriate rules.

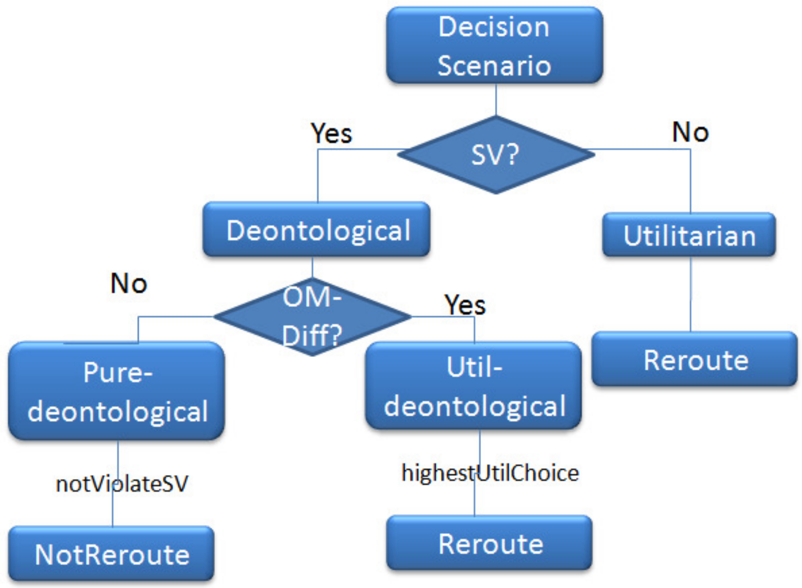


Figure 9: Reasoning in the FPR module

Motivated by moral decision making research, FPR makes decisions based upon protected values, orders of magnitude relationship between utilities and action vs. inaction. FPR can operate in two different modes, calculation-based and recognition based (rule-based and role-based), and it can apply two types of reasoning, utilitarian and deontological. The utilitarian type, which applies the rules of CBA and selects the choice with the highest utility, is invoked when the choice either does not involve a sacred value or there is an order of magnitude difference between the outcome utilities. In situations with protected values and without an order of magnitude difference between outcomes, deontological reasoning is invoked and the choice that does not violate a sacred value is selected.

The FPR module is often used as a bootstrapping method for the AR module. When there are no previously solved cases in memory or when no cases in memory are valid analogs of the scenario the system is trying to solve, the FPR module can prove to be especially useful. Once the FPR module solves some cases, the AR module can take over and use those solved cases as basis for making future decisions in that context. I will discuss this feature in more detail in the experiments section of this chapter.

The highest level rule called by FPR is `makeDecision` which in turn calls `utilCalculation` and `decisionMaker`. `utilCalculation` returns the result of the OMR module which is the qualitative relationship between the utility of the choices. `decisionMaker` fires two sets of rules simultaneously: `utilitarianChoice` and `deontologicalChoice`. These methods are mutually exclusive, returning at most one choice per scenario. If there are no protected values involved in the scenario, `deontologicalChoice` fails and `utilitarianChoice` returns the choice with the highest utility. If there are sacred values involved, then `utilitarianChoice` fails and `deontologicalChoice` returns the results of one of the following rules: `utilDeontologicalChoice` or `pureDeontologicalChoice`. Again these two methods are mutually exclusive and only one can come up with an answer for any given scenario. If there is an order of magnitude difference between the utility of choices, `utilDeontologicalChoice` returns the option with the highest utility. Otherwise, the choice which does not violate the psychological findings on moral decision making discussed previously is returned by `pureDeontologicalChoice`. For more details about the FPR rules please refer to Appendix B.

In the starvation scenario, there is a protected value, people dying, and no order magnitude difference between the utility of the two choices. Therefore, FPR uses the deontological reasoning to select the inaction choice. Figure 10 illustrates the high-level rules used to solve this scenario. Given the breadth of moral reasoning scenarios, the rules implementing FPR are not

```
  (<== (makeDecision ?choice)
        (utilCalculation)
        (decisionMaker ?choice))

  (<== (decisionMaker ?choice)
        (deontologicalChoice ?choice))

  (<== (deontologicalChoice ?choice)
        (pureDeontologicalChoice ?choice))

  (<== (pureDeontologicalChoice ?choice)
         (choices ?decision ?choice)
         (involvesSacredValue ?choice)
         (notOrdersOfMagnitudeDifferent ?decision)
         (isInaction ?choice))

  (<== (involvesSacredValue ?choice)
        (causes-PropSit (chosenItem ?select ?choice) ?consequence)
        (isa ?consequence ?typeOfConcesequnce)
        (relationInstanceMember objectOfStateChange ?consequence ?y)
        (relationMemberInstance isa ?y ?typeOfY)
        (SacredValue ?typeOfY ?typeOfConcesequnce))
(SacredValue Person Dying)
```

Figure 10: Five high level rules and a fact in the KB used by FPR to solve the starvation scenario

complete. Therefore, FPR necessarily fails on some scenarios. These cases highlight the need for the integrated-reasoning approach taken in MoralDM.

3.2.4. Analogical Reasoning Module

As discussed in Chapter 2, analogy plays important roles in decision making. Decision makers frequently use past experiences and draw inferences from their previous choices (Markman & Medin, 2002). Research on the use of analogy in decision making suggests that the comparison between a target and a base involves an alignment process, where structural relations are weighted more heavily than surface similarities (Gentner & Markman, 1997).

To model analogy in decision making, I use the Structure-Mapping Engine (SME) (Falkenhainer, Forbus, & Gentner, 1989; Forbus & Oblinger, 1990; Forbus, Ferguson, & Gentner, 1994), a computational model of similarity and analogy based on Gentner's (1983) structure mapping theory of analogy in humans. There is evidence suggesting that processes governed by the laws of structure mapping are ubiquitous in human cognition (Gentner & Markman, 1997). Moreover, the cognitive plausibility of SME has been examined extensively and in a wide range of experiments (Forbus, Usher, & Tomai, 2005; Gentner & Markman, 1997; Gentner, Rattermann, & Forbus, 1993).

SME operates over structured representations, consisting of *entities*, *attributes* of entities and *relations*. There are both first-order relations between entities and higher-order relations between statements. Given two descriptions, a *base case* and a *target case*, SME aligns their common structure to find a mapping between the cases. This mapping consists of a set of correspondences between the entities and expressions of the two cases. SME produces mappings that maximize *systematicity*; i.e., it prefers mappings with higher-order relations and nested relational structure. The *structural evaluation score* of a mapping is a numerical measure of similarity between the base and the target. It is calculated by assigning an initial score to each correspondence and then allowing scores for correspondences between relations to trickle down to the correspondences between their arguments. These local scores are used to guide the process of constructing mappings, so that mappings with deeper structures are preferred. SME identifies elements in the base that fail to map to the target and uses the common relational structure to calculate *candidate inferences* by filling in missing structures in target. Candidate inferences represent potential new knowledge about the target case that have been calculated from the base case and the mapping. When an expression in the base does not correspond to anything in the target, but the expression is connected to structure in the base that does correspond to structure in the target, a candidate inference is constructed.

Running concurrently with the FPR module, the AR module uses comparisons between new cases and previously solved cases to suggest decisions. When faced with a decision scenario, AR first builds a case using the predicate calculus of the decision scenario and the results of the OMR module. Next, this case is compared using SME with every previously solved scenario in its memory[3]. The similarity score between the new case and each solved scenario is calculated by normalizing the structural evaluation score against the size of the scenario. A previously solved case has to satisfy a number of constraints ("if-then" rules as stated by Weber 1998) before it can get accepted as a valid analog. First, the similarity score between the two cases need to be higher than a certain threshold. This threshold was determined by performing a sensitivity analysis test. For this purpose, four moral decision making scenarios from the test set of my experiments were randomly chosen. Then I varied the threshold of the AR module from 0.6 to 0.35 in 0.05 intervals and measured the performance of the AR module within each of these thresholds. Figure 11 illustrates the performance of the AR as I varied the threshold. The optimal performance was reached at 0.45. The threshold of the AR module was then set to 0.45 and remained constant throughout all the experiments.

[3] Running SME on every one of these cases is inefficient and not cognitively plausible. I plan to incorporate a cognitively plausible model of similarity-based retrieval, MAC/FAC (Forbus, Gentner, & Law, 1995), to make this process scalable. See Chapter 6 for details.

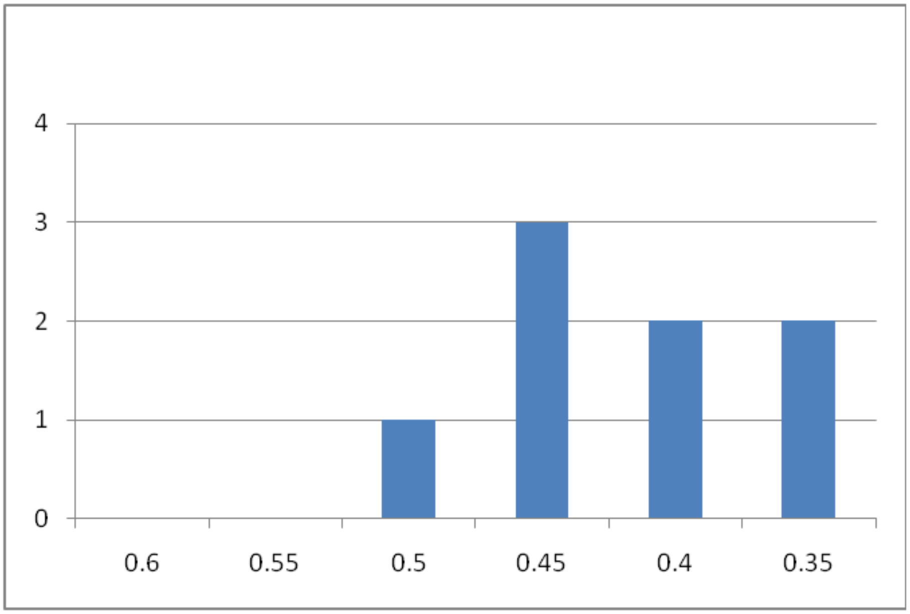


Figure 11: The performance of the AR module as its threshold is varied

Second, both scenarios need to contain the same order of magnitude relationships between outcome utilities. If the scenarios have different order of magnitude relationships, it is likely that a different type of reasoning should be used for the target scenario.

Third, the inferred moral value of the intention for the choice to made should match that of the base. Intentionality here is concerned with motivation for action combined with the willingness to bring about side-effects. In the base case, intention can be inferred from the choice that was made. This inference is performed automatically by EA NLU and the representation for the intention gets included in the base (Tomai, 2009a). If the intention in the base implies a moral value, then the implied intention in the target should also be a moral act. For this purpose the `hasMoralValue` rule is called with its argument being the inferred decision in the target. This rule in turn calls `moralityToBeEvaluated` which first determines what the inferred intention of the agent is for making that choice, by calling `intentionIs`, and then it determines whether that intention is described to have a moral value. Next, the system checks for several other conditions for the inferred intention and its side-effects including whether or not these side-effects are indeed achievable outcomes. The third requirement is especially important for the last experiment of Chapter 4 where the moral value of actions are dependent on the intention of the agent. For more details about the rules and the conditions which an inferred intention has to satisfy please refer to Appendix B.

If these three constraints are satisfied, then the candidate inference indicating which choice to select is considered a valid analogical decision. Otherwise, AR rejects the candidate inference. After comparing against all solved scenarios, AR selects the choice with the highest number confirming analogical decisions. In the case of a tie, AR selects the choice supported by the cases with the highest average similarity score. Because alignment is based upon similarities in structure, similar causal structures and/or sacred values align similar decisions. Therefore, the more structurally similar the scenarios are, the more likely the analogical decision is going to be the correct moral one.

Returning to our starvation scenario example, AR can solve this decision problem through an analogy with the scenario given below, in which the system chose to not transfer funds:

> Your office provides financial assistance to a plant employing 50 workers. If you withdraw this support (which will put 50 workers out of work) you can use the funds to support another plant, which employs 500 workers. Without government support, this second plant will close down.

The analogical decision is determined by the candidate inferences where the decision in the base, inaction, is mapped to the choice in the target representing inaction. Because the transfer of funds scenario contains the same the order of magnitude relationship (almost equal) as the starvation scenario and there is no difference in intention of the agents in the two cases, the system accepts the analogical decision. The base and the target representations, the mapping between them and the candidate inferences resulting from the mapping are included in Appendix C.

3.3. Evaluation

MoralDM was evaluated using moral decision making scenarios from psychology studies. The first experiment includes eight decision making scenarios, each describing two possible outcomes, from Waldmann and Dieterich's (2007) and Ritov and Baron (1999) experiments. The second experiment contains 12 scenarios from Ritov and Baron's (1999) experiments. In all these decision making scenarios, CBA fails to predict the participants' responses. Experiments 1 and 2 evaluate MoralDM as a model for moral decision making and illustrate the importance of using both analogical and first-principles reasoning. In these two experiments, there are cases where one of the reasoning modules fails, but MoralDM is still able to give the correct decision by using the other module. Experiment 3 investigates the claim that as the system accumulates experience the performance of the analogical reasoning module improves. Moreover, I use the result of this experiment to support a theoretical claim that an agent's reliance on recognition-based mode of decision making increases as it becomes more experienced in a domain. In each experiment, I compare MoralDM's decisions to participants' responses as reported by the authors of the original study. When MoralDM's decision matches those of the majority of participants, I consider it a correct choice. In section 3.4, I argue that MoralDM can account for cultural differences in moral decision making. This discussion is continued in the last section of

Chapter 4, where I describe the last MoralDM experiment in which the modeling loop gets closed.

Importantly, the AR module requires previously solved decision cases to draw analogs from. As a result, the AR module always fails on the first case of each experiment because it does not have any cases in its case library to reason with. However, after the FPR module solves the case, it is saved in the case library with a worked solution for future reasoning. Therefore, the AR module can use that case in the next round to suggest decisions.

3.3.1. Experiment 1

I evaluated MoralDM by running it on 8 moral decision making scenarios taken from two psychology studies (Ritov & Baron, 1999; Waldmann & Dieterich, 2007). In all the scenarios used, traditional utility theories fail to predict participants' responses, because they select the choice which provides a smaller overall outcome utility.

For each case, EA NLU semi-automatically translated the simplified English version of the original psychology scenario into predicate calculus. The simplified English and the predicate calculus of two of these scenarios are included in Appendix A. The protected values and the relevant contextual factors are computed via rules. Then the order of magnitude reasoning module calculates the relative relation between the utilities. This relation and the sacred values involved in the case are then sent to the first-principles and analogical reasoning modules. Correct decisions are then added to MoralDM's experiences.

Table 1 displays the results of the first experiment. The analogical reasoning module failed to choose the correct decision in three cases. As discussed above, this module fails on the first case

	# of correct decisions
MoralDM	8 (100%)
First-principles	8 (100%)
Analogical Reasoning	5 (62%)

Table 1: MoralDM results for experiment 1

because it does not have any cases in its memory to reason from. The other two cases involved scenarios for which no close analog could be found due to their considerably different causal structure. For example, in the torpedo scenario (where a torpedo is going to kill 6 soldiers and destroying it will kill 3 other soldiers in another location), unlike the starvation scenario, intervening on the main agent (the torpedo) will cause the soldiers in the first location to survive. These causal differences yield invalid mappings and in turn incorrect decisions. In all three cases, the first-principles module made the correct decision. Overall, MoralDM made the correct choice in all of the scenarios.

3.3.2. Experiment 2

One of the more difficult aspects in building the first-principles reasoning module is the number of rules required to handle the broad range of situations covered in moral decision making. This experiment is designed to test the hypothesis that the analogical reasoning module is capable of making moral decisions in situations when gaps in the knowledge base or rule set prevent the first-principles reasoning module from making a decision. In this experiment, all 12 moral decision making scenarios from Ritov and Baron's (1999) first experiment were used as inputs.

	# of correct decisions
MoralDM	11 (92%)
First-principles	8 (69%)
Analogical Reasoning	11 (92%)

Table 2: MoralDM results for Experiment 2

Unlike the other experiments, 8 could not be translated by EA NLU, so they were encoded manually.

Table 2 displays the results of MoralDM, broken down by reasoning modules. Overall, the system made the correct choice in 11 cases ($p < 0.01$). In 8 scenarios, both modules came up with the correct answer. In three scenarios, the first-principles reasoning module failed to make a prediction, but the analogical reasoning module provided the correct answer. In one scenario, both modules failed.

Examining these failures is instructive. The first-principles reasoning module fails in four of the scenarios because MoralDM's current rules for handling cases with unique structure or content is limited. For example, there is a scenario about Israeli settlements where the first-principles module fails. The system does not have the necessary rules to determine that Israeli land is considered as a sacred value for Israelis, and it cannot be traded off. However, the analogical reasoning module was still able to make decisions in three of these cases based upon similarities with other scenarios, e.g. with a scenario where saving a nature preserve was a protected value.

3.3.3. Experiment 3

This experiment addresses the question of how effective the analogical reasoning module is at learning from experience. Moreover, it simulates how the type of decisions the system makes changes as the number of moral decisions in its KB increases. I measure how performance is affected as a function of the number of previously solved cases in memory. Given the 8 solved scenarios from experiment 1, I created case libraries of every combination of these scenarios. This provided us with 254 different case libraries (8 of size 1, 28 of size 2, 56 of size 3…). Then, with each case library, I tested the analogical reasoning module by running it on each of the scenarios not in the case library. So for each of the 8 libraries of size 1, the test consisted of 7 decision scenarios for a total of 56 decision scenarios.

Figure 12 shows the performance of the analogical reasoning module as a function of the number of available cases. There is a monotonic increase in the number of correct answers as the size of the library increases (r = 0.97, p < .0001). Also, there is a significant decrease in the number of cases where the analogical reasoning module does not come up with an answer (r = -0.95, p < .001). The number of incorrect decisions changes insignificantly from 18% to 25% (r = 0.53, p < 0.22). The statistics reported have been computed by comparing each series against the size of the case library.

The result of this experiment illustrates how the system makes more moral choices as the number of moral scenarios in its memory increases. In other words, it demonstrates the importance of learning from stories and experience in the domain of moral decision making. Without exposure to any moral scenarios, given that the system's rules can only cover a limited number of cases, in most scenarios the system relies on its calculation based mode of decision

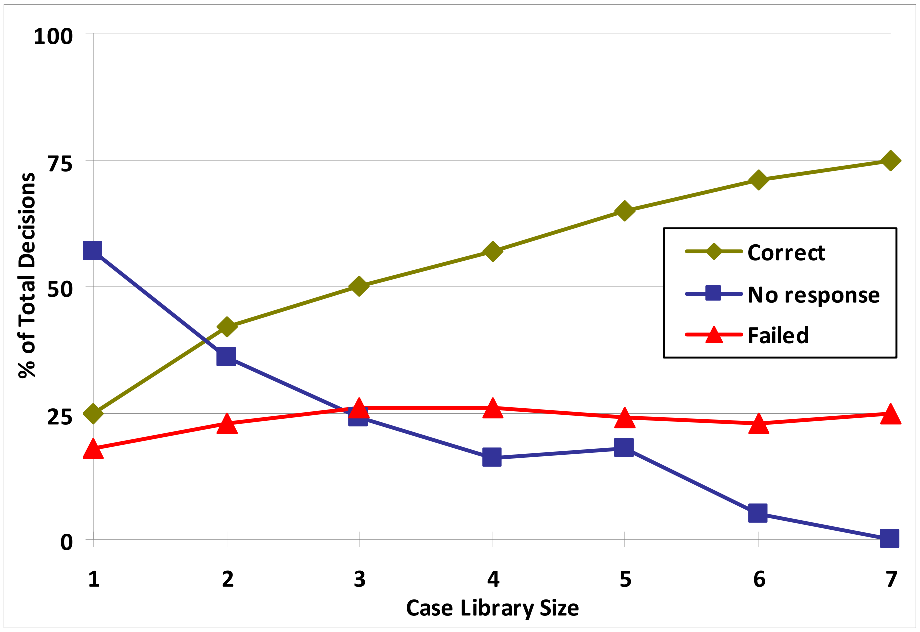


Figure 12: Analogical reasoning results

making. However, using analogy, as the number of scenarios provided to the system increases, the system's reliance on recognition-based mode of decision making also goes up. As a result, with more moral examples in its KB, the system makes more moral choices. In conclusion, this experiment shows how learning and experiencing more moral episodes shifts the mode of decision making of the system from calculation-based to recognition-based.

3.3.4. Modeling Cross Cultural Differences in Moral Decision Making

In the last chapter, I discussed some cross-cultural differences in decision making. For a cognitive model of moral decision making to be successful, it needs to model these differences. One of the most important features of MoralDM is that it not only can be used to explore cross-cultural differences, the decision making process of the system can be trained and/or modified to match that of particular cultures. In this section, I explore some of the ways in which cross-cultural modeling can be achieved using MoralDM.

First, let's explore the role of protected values. Some cross-cultural differences in moral decision making are due to the different protected values that cultures have. For example, nature is considered sacred for Menominee Native Americans (Bang, Townsend, Unsworth, & Medin, 2005). However, it does not have the same value for European Americans living in the same region as the Menominee. We can model these differences in MoralDM by specifying in the knowledge base of the system which protected values the culture we are trying to model has. For example, by specifying in the KB that protecting nature is a protected value for the Menominee, I can capture some of the differences in decision making about nature scenarios for that culture.

Second, as discussed in the next chapter, access to different cultural narratives could lead in to cross-cultural differences in decision making. Certain elements of moral reasoning can be best learned and transferred in narratives, as they are not common situations encountered in daily life. I believe that the impact of cultural narratives on decision making can be captured using analogy. Therefore, due to the reliance of MoralDM on analogical reasoning, by adding story libraries of a certain culture to the KB of MoralDM, we can capture some specific decision making norms of that culture. I will explore this feature of MoralDM in the last section of chapter 4, where I use Iranian cultural stories to model new psychological data.

Third, there are moral rules and norms that are highly salient in some cultures but not in others. These rules articulate modes of decision making and can be seen as adaptation to different environments (Bennis, Medin, & Bartels, in press). Once these unique rules and types of decisions are discovered for a culture, they can easily be added to MoralDM. In this section, I investigate the presence of one of these rules in a particular culture and show how without any changes to the architecture of MoralDM, the FPR module can be easily extended to account for that rule and the type of decision making resulting from it.

I ran the 4 moral decision making dilemmas from Ritov and Baron (1999) which were used in Experiment 1 in Iran. These dilemmas include the starvation scenario which I talked about earlier, a scenario about saving some species of fish by causing some other species of fish go extinct, a scenario about trading old-growth forest with a similar but larger piece of land and one about withdrawing financial assistance from a factory to help employees of a larger factory.

270 participants in Iran completed my questionnaire (male/female 143/127). The participants were either students at University of Tehran or enrolled in the college preparation course (4th year of high school). My assessment (using the same methodology as Ritov and Baron) shows that the majority of my participants did not have sacred values in the fish (42%, χ^2=5.29, $p <$ 0.05), forest (29%, χ^2=38.72, $p < 0.001$) or the financial assistance (23%, χ^2=83.91, $p < 0.001$) scenarios. However, the majority of participants considered human life to be a sacred value (67%, χ^2=34.68, $p < 0.001$). In the forest and financial aid scenarios, my results follow the major finding of Ritov and Baron: people who have sacred values prefer inaction in situations where action violates sacred values. In these two scenarios, the majority of Iranians with sacred values chose inaction (Forest: 89%, χ^2=38.62, p < 0.001, Financial assistance 85%, χ2=26.75, p < 0.001). However, the most striking difference between the results of Ritov and Baron and my experiment in Iran are the results of the starvation and the fish scenarios: the majority of Iranians who had sacred values in these two scenarios, picked the choice representing action (starvation: 79%, χ^2=54.211, $p < 0.001$, fish: 81%, χ^2=36.96, p < 0.001).

I believe this finding is an indication of a different type of reasoning among Iranian participants. When asked about the reason for their choice, both in the starvation and the fish scenario, many participants with protected values indicated that it is justified to sacrifice a smaller group of lives for saving the lives of a larger group. This finding suggests a different decision making strategy among Iranians, a strategy in which the proper action is to sacrifice in order to save a larger group of living things. The idea of sacrifice is embedded in cultural products of many cultures with great saliency in some cultures—in particular, the Iranian culture that concerns us here. I believe it is important to distinguish between a utilitarian choice and a sacrificial choice, as the reasons behind making these types of choices are quite different. In the first mode, participants do not hold the value in question as a protected value and they try to maximize utility when making the choice. On the other hand, in the second mode, the value under consideration is sacred, yet the participants choose to sacrifice that value in order to save and maintain other sacred values. The majority of the participants reasoned in the CBA mode in the forest and the financial assistance scenarios. Also, in the fish scenario given that the majority of the participants did not hold the value in question protected, they reasoned in the CBA mode. However, the sharp distinction in their belief about the sacredness of life compared to values in the other scenarios, as well as their explanation for their choices, indicate that for modeling moral decision making in the Iranian cultural (and in cultures for which sacrifice has great saliency) another type of decision making needs to be added to MoralDM. This new type is referred to as the sacrificial-deontological mode.

In order to model this new type of reasoning, I added a set of new rules to the FPR module under the deontological mode. After calculating the utility of each choice, if the system is operating under the deontological mode (in the presence of protected values), it now checks to see if any of the choices involve sacrificing a value for the greater good of some other values. Next, the system checks to see if sacrificing that value is considered a moral action for the

```
(<== (sacrificialDeontologicalChoice ?choice)
             (choices ?decision ?choice)
             (involvesSacredValue ?choice)
             (involvesSacrifice ?choice ?agent)
             (consideredMoralAct Sacrifice ?agent))

(<== (involvesSacrifice ?choice ?groupEffected)
             (choicesAndConsequences ?choice ?consequence)
             (highestUtilChoice ?choice)
             (typeOfGroupEffected ?consequence ?groupEffected))

(consideredMoralAct Sacrifice Person)
(SacredValue Person Dying)
```

Figure 13: High level rules implemented in the FPR to model the sacrificial-deontological mode

culture the system is modeling. If so, the FPR module returns the choice involving sacrifice. Figure 13 shows the high-level logic used for implementing the sacrificial-deontological mode. After the addition of the sacrificial deontological mode to the FPR module, MoralDM made decisions correctly matching those of the Iranian participants in all 4 cases.

3.4. Discussion

The results of these experiments are very encouraging. As shown in Experiments 1 and 2, my system matches human behavior on a set of decision making scenarios. This result would not be possible without the integrated approach. First, the input in the first and the third experiment was given in natural language requiring EA NLU. Second, these cases all involved sacred values; therefore the order of magnitude reasoning module's computed relationship between outcome utilities is essential to providing the correct answer. Third, the first-principles and analogical reasoning modules were both needed to select the appropriate action.

Due to the breadth of moral decision making, this domain provides an important area to explore the benefits of integrated reasoning. Without both analogical and first principles reasoning, MoralDM would have failed on a considerable number of problems from the first two experiments. In Experiment 1, I demonstrated the necessity of the first-principles reasoning module where there are insufficient appropriate prior cases for analogical reasoning. The analogical reasoning module alone could not have correctly answered the 8 cases. In Experiment 2, I demonstrated that the analogical reasoning module enables the system to handle a wider range of decision making scenarios where gaps in the knowledge base and/or rule sets prevent the first-principles reasoning module from answering correctly. Without the analogical reasoning module, MoralDM would have failed on three more cases.

The results of the first two experiments emphasize the importance of integrated reasoning. Given that the AR module fails if there are no solved cases in the memory, the FPR module was used to bootstrap the AR module. Moreover, in the second experiment when the gaps in the rules caused the FPR module to fail, the AR module took over and solved the cases.

Experiment 3 provided additional support for the importance of the AR module within MoralDM and demonstrated how it serves as a learning component. The results show a significant improvement in the analogical reasoning module's performance as the number of cases in MoralDM's memory increased. This result provided the impetus for further empirical experiments to analyze the role of analogy in moral decision making. The next chapter of this thesis is dedicated for this analysis.

In section 3.4, I explored the ways in which MoralDM can be used to capture and explore cross-cultural differences in moral decision making. First, by adding protected values of a culture, we can predict how that people from that culture reason about cases involving those values. Second, we can add a library of important narratives of a culture to the KB of MoralDM and use the AR module to reason about novel cases which resemble the narratives in structure.

Third, the system can be informed about rules and types of decision making which might be more salient in some cultures than others. Given how the rules in the FPR module are organized, adding and subtracting rules and types of decision making is straightforward for experimenters who are familiar with formal logic. When the system solves a dilemma using the added rules, the AR module can take over and use the logic used in the solved scenario as a basis for solving similar scenarios.

The preliminary results of this work were originally published in (Dehghani, Tomai, Forbus, & Klenk, 2008; Dehghani, Tomai, Forbus, Iliev, & Klenk, 2008a; Dehghani, Tomai, Forbus, & Klenk, 2008b)

3.5. Related Work

Reasoning with orders of magnitude is a form of commonsense reasoning. Previous research has identified its value in situations when complete quantitative information is not available or when tackling problems involving complex physical systems. Order of magnitude reasoning has been used in several engineering tasks (Dague, 1994; Mavrovouniotis & Stephanopoulos, 1990; Dague, Deves, & Raiman, 1987). My work is the first to apply this formalism to cognitive modeling.

Computational models of cultural reasoning are receiving increasing attention (e.g. ICCCD[4]). The CARA system (Subrahmanian, et al., 2007) is part of a project to "understand how different cultural groups today make decisions and what factors those decisions are based upon". The major difference between MoralDM and CARA is that MoralDM focuses on the cognitive

[4] http://www.umiacs.umd.edu/conferences/icccd2008/

plausibility of the model while CARA does not aim to be a cognitive model. CARA highlights the need for integrated systems in tackling these complex problems by incorporating semantic web technologies, opinion extraction from weblogs to build cultural decision models consisting of qualitative rules and utility evaluation. While I agree that qualitative reasoning must be integrated with traditional utility evaluation models, I also believe that analogy plays a key role in understanding and modeling cultural reasoning. Moreover, my approach differs as I evaluate my system against psychological studies, which helps ensure its judgments will be like those that people make.

With the goal of building ethical reasoning machines, there have been a number of research projects building ethical advisors. The MedEthEx system (Anderson, Anderson, & Armen, 2006) uses inductive logic programming (ILP) techniques to learn decision principles from training cases. This system provides advice in the domain of ethical dilemmas by learning relationships between a novel case and known rules inputted to the system. Mclaren's Truth-Teller and The System for Intelligent Retrieval of Operationalized Cases and Codes (SIROCCO) (2006) use case-based reasoning to highlight relevant ethical considerations and arguments to a human user. Truth-Teller compares cases presenting dilemmas about whether or not the agent should tell the truth. SIROCCO deals with the domain of engineering ethics and predicts which stored principles and cases are relevant for a new dilemma. In line with these approaches my analogical reasoning module also uses known solved cases to guide moral decisions. However, unlike MoralDM, these models do not capture the effects of sacred versus secular values on decision making.

Rzepka and Araki (2005) propose an extension to their web-based knowledge discovery system GENTA (General Belief Retrieving Agent) for learning ethical behavior by extracting information from the Web. Their proposed system relies on statistical approaches for extracting opinions, behaviors and consequences relevant to moral issues. Once this extraction is performed, the system categorizes these data according to their positive or negative outcome. Their goal for this project is to implement their techniques in housework robots "to assure safety of the users and the system" (Rzepka & Araki, 2005). I have been unable to verify whether their proposed method was ever implemented in a working system. Guarini (2005) proposes a neural network approach for modeling ethics. His model classifies actions as acceptable or unacceptable depending on the motivations behind making the choices and their consequences. He proposes two artificial neural networks, one feedforward and one recurrent, which can be used as “moral case classifiers”. These networks classify the inputs, a set of variables, given to the system and do not have deep knowledge and understanding about the cases nor the rules. Due to the shallow reasoning that these statistical systems perform, they are unable to produce reasonable explanations for their decisions. Guarini (2005) acknowledges this limitation and argues that the second major limitation of his system is its inability to process arguments, as opposed to processing a set of variables.

Pereira and Saptawijaya (2007) propose a framework for modeling moral reasoning using ACORDA which is an implementation of prospective logic programming. In their model, the only principle used when making moral decisions is the principle of double effect (Mikhail, 2007), which is only one of the five main principles encoded in MoralDM. Given that Pereira and Saptawijaya’s system relies only on one single rule, it is very limited in the type of cases it can solve. While these approaches are attempting to encode expert ethical knowledge, my work attempts to model the moral reasoning of average human subjects leading to a very different methodology. I evaluate the system empirically through comparisons to human data. This is a key element to building agents that can reason about moral decisions.

Deontic logic (Hohfeld, 1913; Wright, 1951; Hilpinen, 2001) is an extension of modal logic used for formalizing ethical principles. This formalism adds special operators to first-order modal logic for permission, obligation and prohibition. For example in this frame work, $\circ P$ is interpreted as *it ought to be the case that P*, where *P* represents some proposition. In other words, deontic logic is the logic of ideals versus actual behaviors used for reasoning about ethical and legal situations (Meyer, Dignum, & Wieringa, 1994). However, since the early days of deontic logic number of paradoxes were discovered in this framework. These paradoxes are logical statements which hold in deontic-logical systems, but are counterintuitive in commonsense reasoning (Åqvist, 1984). For example, in this formalism ought-to-help-Jones-who-is-robbed implies Jones-ought-to-be-robbed (Meyer, Dignum, & Wieringa, 1994). This paradox is known as the Good Samaritan paradox. Recent updates to this formalism have solved some of these paradoxes (Horty, 2001), however still some inconsistencies remain in deontic logic. Bringjord, Arkoudas, and Bello (2006) use Murakami's (2004) axiomatized deontoic logic as the bases of their system for formalizing moral codes about what *agents ought to do*. They propose a decision procedure based on their mechanized multi-agent deontic logic, where a robot would only "take an ethically charged action if it could formally prove that the action is permissible". In this framework, the content in which the rules are to be evaluated is not taken into account and the rules universally apply in all situations. Moreover, Bringjord, Arkoudas, and Bello (2006) assume that the agents will be preprogrammed with deontic logic and therefore the role of learning is overlooked. In the same line of research, Power (2006) proposes a theoretical framework in nonmonotonic logic for modeling Kant's categorical imperatives. However, it is unclear whether this framework has ever been implemented in a working system.

Slade (1994) proposes a program called VOTE which models the decision making process of members of Congress in the United States House of Representatives. This model predicts how these representatives will vote on a particular bill regarding a certain issue. VOTE is a case-base reasoning model which utilizes databases of previously voted-on cases.

My combination of analogical and first-principles reasoning is inspired in part by Winston's (1982) use of both precedents and rules to reason about a situation. His work was hampered by the lack of off-the-shelf large-scale knowledge bases, and the technologies for NLU and analogical reasoning have improved since then.

4. The Role of Analogical Reasoning in Moral Decision Making

4.1. Introduction

Moral behavior is defined by Kuhmerker (1975) as intentional sets of actions with social consequences. These social consequences are evaluated as moral (or immoral) with respect to the values, norms and duties defined within one's culture. Such norms and rules make up the cultural wisdom of societies. Cultural products created over generations are responsible for storing and transmitting this cultural wisdom (Weber & Hsee, 1998). One type of cultural product that may underlie culturally specific moral values is core cultural narratives. MacIntyre (1981) argues that cultural narratives are in essence the "historical memory" of a society and form the moral framework of it. In other words, in order for one to be moral, one has to be familiar with the traditions and stories of one's culture (Lockwood, 1996). Lyotard (1984) believes that grand cultural narratives are the principle way in which a culture legitimates itself, "these narratives define what has the right to be said and done in the culture in question, and since they are themselves a part of that culture, they are legitimated by the simple fact that they do what they do" (p 23). Cultural narratives are often used to inform us about our identity with respect to the culture and the society we live in (Barbour, 1974). Winston (1998) claims, "Our ability to author the moral self is therefore dependent upon our understanding of the virtues embedded in the social roles we are born into and these are learned in part from the stories which are part of our heritage". In a similar line, Prasad (2007) conducted ethnographic research in Sringeri in

Southern India looking at how oral narrative shape moral identities. She says "…narratives, spontaneously shared and ceremonially delivered, hold up the moral self as dynamic and gendered, with a historical presence, a political agency, and a capacity for artistic expression that mediates many sources of knowledge to articulate appropriateness of conduct." In fact, she says that the way in which cultural narratives about morality are interpreted and reinterpreted at every telling are instrumental in the complex nature of moral reasoning. Cultural narratives lay out the principles used in the decision making of the protagonist of the story and the consequences resulting for that decision with respect to the content of the story, which carries values of the culture. Kilpatrick, Wolf and Wolf (1994) argue, "The dramatic nature of stories enables us to 'rehearse' moral decisions, strengthening our solidarity with the good". It is my belief that certain elements of moral reasoning can be best learned and transferred in narratives, as they are not common situations encountered in daily life. Great cultural narratives, such as those contained in most religious texts or in folk stories, can deeply imprint our long term memory, whether or not we ever encounter these situations in real life. It is not implausible to think that those values seep into our being and affect our reasoning. In other words, we can view collections of cultural narratives as moral compasses for cultures, helping us distinguish moral actions from immoral ones.

As discussed in Chapter 2, the link between analogy and decision making has been explored from various perspectives. However, this role has not yet been systematically examined in the domain of moral decision making. The result of the third experiment of the previous chapter illustrated that as an agent learns more moral stories, it can rely more on the recognition-based mode of decision making. This result provided the impetus to further examine the function of cultural narratives and analogical reasoning on moral decision making. In this chapter, I investigate the role of cultural narratives in understanding novel moral situations. I examine whether the processes by which core cultural narratives are applied in people's lives follow the principles of analogical retrieval and mapping. If so, then moral reasoning should manifest the keynote phenomena that characterize analogical processing. In particular, I examine how analogical accessibility and alignability influence the use of canonical moral narratives. I also show how access to different moral stories results in differences in moral preference across cultures. I report on the results of a series of experiments performed among Iranian and American participants. My results indicate that analogical accessibility to cultural narratives that are similar in structure to a given dilemma is the differentiating factor in my participants' responses across the different variants and between the two cultural groups.

This chapter is organized as follows. I begin by summarizing relevant results on the role of narratives in moral education. Next, I discuss the role of analogical reasoning in Judeo-Christian and Islamic jurisprudence. Then, I discuss the role of similarity in long-term memory retrieval and inference. Next, I explain my hypotheses and describe my experiments and results. Then, I use MoralDM to simulate the results of the experiments of this chapter.

4.2. The Role of Narratives in Moral Education

For thousands of years, narratives have been the prominent method for passing moral values from a generation to the next. Major religious traditions and holy texts can be viewed as collections of moral stories and narratives which embody the codes of conduct that the followers of that religion are to abide by. For example, the Qur'an repeatedly refers to its content as a set of narratives: "We narrate to you (O' Prophet) the most excellent of the narratives by (means of) what We have revealed to you this Qur'an” (The Qur’an 12:3). These narratives, carrying moral rules and codes of conduct, encourage the followers of that religion to live by the principles exemplified in the stories. In other words, narratives discussed in the holy texts layout the foundations of the religion itself. Kilpatrick (1992) stresses the important relationship between stories and religion, “Most cultures have recognized that morality, religion, story, and myth are bound together in some vital way, and that to sever the connections among them leaves us not with strong and independent ethical principles but with weak and unprotected ones”.

The approach of teaching moral values through cultural narratives, sometime referred to as the Great Tradition, has its advocates and opponents. Laying out the full detail of the arguments of these two camps is beyond the scope of this thesis. However, a crude classification of these groups would distinguish two different schools of philosophy: one which advocates *Kohlbergian* approaches to moral development and one which does not.

Kohlberg’s model (1981; 1984) assumes that moral reasoning is the product of the development of abstract moral principles, grounded on objective, ideal and impersonal propositions. Vitz (1990) summarizes Kohlberg’s theory, “In short, the model presents moral development as a process of abstract cognitive development, as a growth in rational competence expressed in increasingly sophisticated principles of moral reasoning”. Kohlberg’s theory can be thought of as expansion of Piaget’s (1932/1965) theory of mental development. In Kohlberg’s view, children’s moral judgments go through six developmental stages, starting from heteronomy to increasing autonomy. Presenting a complete picture of Kohlberg’s theory is again beyond the scope of this thesis. I will only briefly summarize some relevant research.

The Kohlbergian and Piagetian cognitive developmental theories of learning in moral growth, which assume that morality constitutes a set of principles valid for every culture, have been seriously challenged by Gilligan (1982), MacIntyre (1981) and their colleagues. These perspectives claim that the moral self is developed and evaluated with respect to historical and social norms. Gilligan and her followers see culture and language, and in general contextuality and meaning, as the fundamental constituent of moral meaning and place "*narrative* at the center of moral life" (Winston, 1998). Gilligan argues that the narrative structure is best suited to hold and express moral knowledge, and it is through these narratives and stories that our moral propensities develop.

MacIntyre (1981) has a similar view to Gilligan and critiques the Enlightenment Project of Immanuel Kant (1781/2003), in which the basis of morality is pure reason and universal axioms. He places "narrative at the heart of his concept of the unity of the moral life and at how morality is learned" (Winston, 1998). In his view, Aristotelian virtues are essentially only comprehensible through narratives, as narratives embody historical and cultural aspects of a society. He argues, "There is no way to give us an understanding of any society, including our own, except through the stock of stories which constitute its initial dramatic resources" (p 216). Moreover, he stresses the great power of stories in shaping children's lives as he famously argues "Deprive children of stories and you leave them unscripted, anxious stutterers in their actions as in their words" (p 216). Kilpatrick (1992) argues that stories and narratives give children a common reference point by anchoring them in their culture and providing them with collections of moral examples to follow. Moreover, he hints that stories maybe the source of sacred values, "Our 'sacred' memories may find their source in stories".

Bruner (1986) classifies the spectrum of human thinking into two qualitatively different modes: propositional thinking which consists of logical argumentation and abstract rules, and narrative thinking which is essentially a description of reality, requires imagination, and presents human and interpersonal situations happening in a particular point in space and time. Vitz (1990) argues that given that our understanding of moral issues is an "interpersonal, emotional, imagistic, and story-like phenomenon" it fits well with the narrative mode of thought and Kohlberg's model which classifies moral reasoning to abstract moral principles fails to account for these prerequisites of our moral reasoning. In other words, given that people interpret moral dilemmas in the context of personal narratives, our moral thinking cannot fit in the "abstract cognitive principles" of Kohlberg's model. Vitz argues that moral education, therefore, cannot take place with abstract examples which students cannot relate to and claims that "narratives (stories) are a central factor in a person's moral development". Moreover, he claims that one of the only universal aspects of moral education is the use of stories.

Lockwood (1996) argues that another important reason for teaching morality to children through stories is that stories help the content of morality to stay with children. Furthermore, he claims that when thinking about these narratives children can imagine what it would be like to be in that particular context and face the moral dilemma expressed in the narrative. This act of rehearsing the moral context through imagination, "allows them to decide what the best way to act in a similar situation would be" (Lockwood, 1996).

Kohlberg's theory has more recently been reformulated by the *neo-Kohlbergian* camp (Narvaez, 2005) whom have shifted the focus from moral judgment stages to moral judgment schemas. However, even this nuanced theory denies the importance of culture and cultural narratives in moral reasoning (Narvaez, 2002). Below are a set of studies which Narvaez (2002) uses to reject the role of narratives in moral education.

In a study, Wilder (1980) presented a set of moral stories to children and asked them to retell the stories using puppets. He reports that younger and older children show differences in recall with regards to type and amount of story elements, with older children recalling more moral themes and structures while younger children focus more on surface features. In a similar line of research, Johnson and Goldman (1987) presented children with bible stories and gave them a task in which they had to recognize the rules implicated in the stories, and a separate task in which they had to categorize the stories based on the rules involved in them. Younger children tended to categorize the stories based on surface features (e.g. items involved in the story) rather than structural features (e.g. types of rules involved). Narvaez, Gleason, Mitchell and Bentley (1999) first presented participants from 3rd and 5th grade and university students with four moral stories. After reading each story, participants were presented first with a list of alternative stories and were asked to rate the closeness (similarity) of the alternative story to the base moral story. Then they were provided with a list of vignettes, each containing a certain moral theme and they were asked to choose one which resembles the original story the most. The alternative stories and themes provided to the participants were different from the original story with regards to different features. The first group of stories had the same actions, but a different theme and actors from the original story. The second group had the same actors but different actions and themes. The third group involved stories with the same settings but different theme, action and actors. And finally, the fourth group contained stories which had the same theme as the original story but differed in the actors and actions involved. Within these four classes of stories, the one with the same theme and different actors and actions was considered by the experimenters to be the most similar one. Students were also asked to answer true/false questions about the stories for measuring their overall comprehension of them. Their results showed that performance improved with increase in age and there was a significant difference between 3rd and 5th graders even after controlling for reading comprehension. The younger children less frequently chose the target which resembled the original story in structure. Moreover, for all ages the most attractive choice was the choice which had the same actions as the target, but different theme (surface similarity). They conclude from these results that there are developmental differences in moral understanding (Narvaez, Gleason, Mitchell, & Bentley, 1999). Narvaez (2002) questions the assumption that moral stories help students in their moral literacy. She uses the results of the above experiment on moral theme comprehension in children along with her results on the relationship between moral judgment development and reading (Narvaez, Bentley, Gleason, & Samuels, 1998), to conclude that because children cannot extract the correct moral themes from moral stories, these stories cannot be helping them in their moral literacy.

The argument that I present in chapter, which emphasizes the role of analogical reasoning in moral cognition, can potentially explain Narvaez et al.'s experimental results from a different perspective. Specifically, findings on the cognitive processes involved in analogical reasoning and similarity (Gentner, 1983; 1988; Goldstone, Medin, & Gentner, 1991; Markman & Gentner, 1993) in general show that younger children tend to focus more on surface similarity (similarity between attributes or features) rather than relational similarity which require having deep relational representations. However, as children gain knowledge, and in other words build deeper structural representations, they increasingly focus more on relational similarity (Gentner & Rattermann, 1991). This *relational shift,* as Gentner and Rattermann (1991) refer to it, has been used to explain differences in the performance of older and younger children (and adults) in variety of different tasks (Gentner & Rattermann, 1991; Uttal, Gentner, Liu, & Lewis, 2008). Therefore, in my opinion Narvaez's results can even provide further proof for the importance of narratives and analogical reasoning in moral education. Even though, younger children may not fully understand the overall theme of the stories the first time they hear it, through frequent repetition and retrieval of the stories, as well as building deeper relational representations of them as they get older, they will better comprehend the underlying moral theme the stories try to convey.

4.3. Analogical Reasoning in Judeo-Christian and Islamic Jurisprudence

As discussed above, religious books and traditions can be viewed as collections of stories and moral cases which the follower of that religion are encouraged to follow and apply in their daily lives. Moreover, when faced with novel cases religious scholars use these cases to draw inferences from via analogical reasoning. According to the Encyclopedia of Christian Theology (Lacoste, 2004), analogy "designates the gap between human knowledge of God and God himself" (p. 27) and is applied to relations between sensible and the divine, as a path toward "knowledge of the unknowable God".

Hasan (2007) defines the process of analogical reasoning in the Islamic tradition (*qiyas*) as follows: "In a general manner, we may define *qiyas* as comparison of a case not covered by the text with a case covered by the text on account of their common *Shari'ah* value (*'illah*) in order to apply the law of the one to the other." (p. 15) Hasan (1986) argues that *qiyas* provides a "beacon light" to a jurist for exercising *ijtihad* (jurisprudence) on questions not covered by the Qu'ran and the *Sunnah* (the collection of Mohammad's sayings). Hassan argues, "It is unanimously agreed that *qiyas* reveals the law which already exists, it does not originate it... Thus the law is originated by God and discovered by *qiyas*". Reasoning based on the similarity of two parallel cases is widely used in the *Hadith* (the sayings of Mohammad) literature which emphasizes the frequent use of *qiyas* in Islam. According to the rules of Islamic Jurisprudence, each analogy is composed of four parts: The *asl* (the original case), The *far'*(the parallel case not covered by the text), the '*illah* (cause of the textual law of the original case or the "*ratio legis*") and the *hukm al-asl* (the law of the original case covered by the text). For example (from Hassan 1986), drinking *nabidh* (a traditional alcoholic drink made from dates) is not permitted in Islam on the basis that it is similar to wine and drinking wine is forbidden according to the Qur'an. In this case, the original case is "drinking wine is forbidden", the *far'* is drinking *nabidh* for which we are trying to find a rule of law for, the cause (or the *ratio legis*) of the law is intoxication which holds for both cases, and the rule of the original case is the prohibition of drinking wine. Islamic Jurisprudence has put forth several conditions for instances where analogical reasoning cannot be applied. For example analogy is not operative in cases which are only literally or physically similar or when the law of the original case is exceptional.

4.4. Similarity, Retrieval and Alignment

To explore how analogical reasoning influences moral decision making, I conducted a set of studies. In these studies, I varied the kind of similarity between the target given to the participants and the core cultural story (which is never presented). The first question is how similarity between the target story and the core story will influence reminding of the core story. In general, surface similarity is the best predictor for whether a current target story will retrieve a given base story from long term memory (LTM); and structural similarity is the best predictor for inference (Forbus, Gentner, & Law, 1995; Gentner, Rattermann, & Forbus, 1993; Holyoak & Koh, 1987; Ross, 1989). However, structural similarity can also influence retrieval of prior cases. Structural retrievals are more likely among domain experts than among novices (Novick, 1988);

and more likely among learners who have previously compared the base story to another analogous story (Gick & Holyoak, 1983; Gentner, Loewenstein, & Thompson, 2003) (Of course, these phenomena may be related).

Because cultural narratives are deeply entrenched as part of oral culture, participants are likely to have heard and compared various versions, resulting in a somewhat schematized encoding (see Gick & Holyoak, 1983; Gentner, et al. in press). Therefore the retrieval rate of these stories may be relatively less dominated by surface similarity than is typically found in experimental situations (Blanchette & Dunbar, 2000). Thus the question for retrieval is (a) whether participants who are familiar with these stories will show remindings to the core cultural story; and, if so, (b) whether their reminding will be influenced by surface similarity, structural similarity, or both.

The second set of questions and predictions concern inference. Assuming that the core narrative is accessed, in order to draw inferences, it must first be aligned with the target story (Colhoun & Gentner, 2009; Clement & Gentner, 1991; Markman, 1997). The correspondences created by this alignment are used to import knowledge from the base representation into the target. Thus, if analogy is operative, then participants who are familiar with the base stories should make more inferences from the core narrative for targets that are structurally alignable with the core narrative.

4.5. Experiments

In this section, I present a set of experiments which investigate whether the processes by which cultural narratives are applied in people's lives follow the principles of analogical retrieval and mapping. In sum, my chief prediction is that, for participants who know the stories, moral reasoning should abide by the key constraints of analogical processing: that is, structural similarity to the core narratives should guide inference. Of course, I predict no such pattern for participants who are not familiar with the stories. With respect to retrieval, the question is whether participants familiar with the stories will show the typical pattern (that is, surface similarity as the main predictor of retrieval), or whether they will show the pattern characteristic of experts (of structural similarity also as a strong predictor of retrieval).

In order to compile a list of salient stories for a given culture, I performed an Internet-based pilot study using 199 Iranian participants. Among other questions, participants were asked to list the top 10 cultural and moral stories they could think of. Based on participants' answers, I compiled a list containing the most referred to non-religious and non-political narratives. Interestingly, all the stories in this list focused on the high moral value of sacrifice. The idea of sacrifice is embedded in narratives of many cultures, with great saliency in some cultures—in particular, the Iranian culture that concerns us here.

Next, for each of these narratives, I developed four different variants: surface changes relative to the base scenario; structural changes; both surface and structural changes; and changes that affect the core cultural values (sacred values) that underlie the narrative. In the latter case, the prediction was that an alteration of the core sacred values should decrease structural similarity. In all variations, I tried to leave the choice of action unchanged, and only vary the intention of the agents or the information provided in the scenario. A key feature of these studies is that the base domain (the cultural narrative) is never presented to participants. I am predicting that such narratives are sufficiently entrenched in the minds of members of the culture that no presentation is necessary.

My hypotheses are that for Iranians, (1) changing the surface structure of the scenarios should still allow inference from the original cultural stories, while changing the deep structure should block the inference; (2) the rate of retrieval of cultural narratives should vary based upon the degree of surface similarity and also (because of schematization) structural similarity with the new scenario; (3) Americans, who lack these cultural narratives, should show no difference between these variations. In conclusion, I suggest that by using analogy we apply a moral theme, a certain relational structure from one domain (that of the cultural narrative) to a novel, but structurally similar domain.

4.5.1. First Narrative: Pourya Vali

To test these hypotheses, I created story variants for the following cultural narrative, prominent in Iranian culture:

Base Story:

Pourya Vali was the most famous wrestler of his time. The morning before wrestling with a young athlete from another province, he goes to a mosque and sees the mother of the young athlete praying and saying “God, my son is going to wrestle with Pourya Vali. Please watch over him and help him win the match so he can use the prize money to buy a house”. Pourya Vali thinks to himself that the young wrestler needs the money more than he does, and also winning the match will break the heart of the old mother. He has two choices, he can either win the match and keep his status as the best wrestler in the world or he could lose the match and make the old mother happy. Even though he was known not to ever lose a match, he loses that one on purpose.

Surface change (ΔSF):

Ali is the greatest ping pong player of his city. The morning before a match with a young athlete from another city, he goes for a walk outside the stadium and sees the mother of the young athlete praying and saying “God, my son is going to play a match with Ali the famous ping pong player. Please watch over him and help him win the match so he can use the prize money to get married”. Ali has two choices, he can either win the match and keep his status as the best ping pong player or he could lose the match and make the old mother happy.

Structure change (ΔST):

Ali was the most famous wrestler of his city. The morning before wrestling with a young athlete from another province, he goes to a mosque and sees the mother of the young athlete praying and saying “God, my son is going to wrestle with Ali. Please watch over him and help him win the match so he can use the prize to buy me new expensive clothes”. Ali has two choices, he can either win the match and keep his status as the best wrestler in the world or he could lose the match and make the old mother happy.

Surface + Structure change (ΔSS):

Ali is the greatest ping pong player of his city. The morning before a match with a young athlete from another city, he goes for a walk outside the stadium and sees the mother of the young athlete praying and saying “God, my son is going to play a match with Ali the famous ping pong player. Please watch over him and help him win the match so he can use the prize to buy me new expensive clothes". Ali has two choices, he can either win the match and keep his status as the best ping pong player or he could lose the match and make the old mother happy.

Sacred Value Change (ΔSV):

Ali was going to wrestle against the most famous wrestler of his city. The morning before the match, he goes to a mosque and sees the mother of the famous athlete praying and saying "God, my son is going to wrestle with young Ali. Please watch over him and help him win the match so he can keep his status as the best wrestler in the world". Ali has two choices, he can either win the match and beat the best wrestler in the world or he could lose the match and make the old mother happy.

After reading one of these dilemmas, the participants were asked the following questions:

1. What should Ali do?
    a. Win the match
    b. Lose the match and make the old woman happy
2. What narrative does this scenario remind you of?
3. If it reminds you of any narratives, please list the similarities between the two.
4. Please list the differences between the two.

Choice 'a' in question 1 corresponds to the utilitarian choice, that is the choice that brings the highest overall utility to the agent. Choice 'b' represents the choice involving sacrifice, where the agent disregards his own immediate utility for the betterment of others (deontological choice). The American group received English translations of the above scenarios with the changes in the names, sports and the locations such that they would be more familiar to American audiences (e.g., Andrew instead of Ali, tennis instead of wrestling, etc.). Translations of the stories and their variants were done by independent translators. For the Farsi version of all the stories used in this experiment, their variants and the versions used for American participants please refer to Appendix D.

4.5.1.1. Method

364 participants in Iran (mean age = 18.67; Female/Male: 191/173) completed my questionnaire. These participants were either students at University of Tehran or enrolled in the college preparation course (4th year of high school). The control group was 48 Northwestern undergraduates (mean age = 18.91; Female/Male: 28/20). Each participant received one target variant (randomized across subjects). For the Iranian participants, the answer to the second question was coded as a recall only when they recalled the cultural narrative. However, for the control group a recall was coded when they indicated any story retrieved from LTM (including children's stories, movie plots, etc.) as the base stories are not known by the Americans. The answers to questions 3 and 4 were coded using the following scheme: if participants reported attribute similarities/differences to/from the base, these were coded as surface similarities/differences, whereas functional/relational similarities/differences were reported as structural similarities/differences.

4.5.1.2. Results

The total and conditional retrieval rates (question 2) are reported in Table 3. The retrieval results show dependence on both surface and structural similarity. The variant with surface changes (ΔSF) led to the best retrieval rate (66%), significantly better than the ΔSV variant (43%) (χ^2= 8.68, df = 1, p < 0.005). The ΔSF variant also led to more retrievals than the ΔSS situation (51%) (χ^2 = 3.65, df = 1, p = 0.05).

Among the control group, four participants recalled a story: Two of the four mentioned pop-culture movies and the other two referred to other stories used in the experiment.

Table 3: Retrieval rates of the core narrative for Iranians in the first narrative

Alternative Chosen	ΔSF	ΔST	ΔSS	ΔSV
Utilitarian	0.11	0.20	0.09	0.10
Sacrificial	0.55	0.36	0.42	0.33
Total Retrieval	0.66	0.56	0.51	0.43

The proportion of sacrificial choices (choice b) to the total number of selected choices for each variant is reported in Table 4. As predicted, Iranians who received the ΔSF target were highly likely to make the sacrifice inference. Those receiving ΔSS were also highly likely to make this inference. There was a significant difference between the following variants: ΔSF and ΔST (χ^2 = 6.53, df = 1, p < 0.01), ΔSF and ΔSV (χ^2 = 4.38, df = 1, p < 0.05), ΔST and ΔSS (χ^2 = 5.81, df =

1, $p < 0.05$) and ΔSS and ΔSV ($\chi^2 = 3.79$, $df = 1$, $p < 0.05$). For the control group, there were no significant differences among the variants[5].

Among Iranian participants who reported structural similarities to the core narrative, a larger number chose the sacrificial choice (92%) than the utilitarian choice (8%) ($\chi^2 = 96.69$, $df = 1$, $p < 0.001$). However, among those reporting structural differences, a larger number chose the utilitarian choice (71%) than the sacrificial choice (29%) ($\chi^2 = 8.64$, $df = 1$, $p < 0.005$). In the other direction, among the Iranians who chose the sacrificial choice, a significantly larger number reported structural similarities to the cultural narrative (48%) than reported surface similarities (23%) ($\chi^2 = 17.30$, $df = 1$, $p < 0.001$). Among Iranians who chose the utilitarian choice, the reverse held: a significantly larger number reported structural differences (54%) rather than surface differences (5%) ($\chi^2 = 18.70$, $df = 1$, $p < 0.001$). Note that even those who chose the utilitarian option still mostly made reference to the cultural narrative.

In the expected direction, among those in the ΔSS condition who chose the sacrificial option, a larger number of alternate stories (31%) were reported than those in the ΔSF condition (12%) ($\chi^2 = 4.52$, $df = 1$, $p < 0.05$). Among the Iranians who were reminded of the core story, a significantly larger number of participants chose the sacrificial choice (76%) than chose the utilitarian choice (20%) ($\chi^2 = 157.53$, $df = 1$, $p < 0.001$).

Logistic regression revealed a significant difference in the trend of answers to these variants between the Iranian participants and the control group ($z = -3.868$, $p < 0.001$).[6]

[5] A power test revealed that even had there been the same number of subjects in the American group as in the Iranian group, the probability that all of the above differences would hold among the Americans would have been very low (less than 2.5% for the first study, 10% for the second and less than 5% for the third experiment).

[6] An ANOVA power test suggests that the difference would remain significant if there were an equal number of subjects in both groups.

Table 4: Proportions of sacrificial choices to total number selected choices for the first narrative

	ΔSF	ΔST	ΔSS	ΔSV
Iranian Group	0.83	0.65	0.82	0.68
Control Group	0.07	0.00	0.00	0.00

4.5.1.3. Discussion

As predicted, Iranians were highly likely to draw the inference suggested by their core narrative, especially when they could align the structure of the target with that of the core narrative. Also as predicted, Americans (who lack this core narrative) showed no such pattern; there were no differences among the variants.

Although the overall pattern for Iranians—particularly for the conditional retrieval rates—strongly suggests that cultural narratives were important in their decisions, there were some puzzling findings. In retrieval, the ΔSF variant led to same number of retrievals of the cultural narrative as ΔST (p = 0.2), contrary to the usual finding that retrieval depends most on surface similarity. In inference, while I found the expected pattern—fewer sacrificial inferences when the structure was changed (ΔST and ΔSV) than when only the surface was changed (ΔSF)—I also found a high rate of sacrificial inferences when *both* structure and surface were changed (ΔSS).

I suspect that this pattern is driven in part by fact that, as mentioned earlier, there are several other cultural stories similar to the Pourya Vali story. These acted as competition during memory retrieval, or might have blended with the Pourya Vali story, so that when both surface and structure were modified (ΔSS), many Iranians were reminded of other stories that have surface and structural resemblance to the ΔSS variant. In fact, many Iranians were reminded of other stories that have surface and structural resemblance to the ΔSS variant (chiefly a moral story about another wrestler and a moral story about a running match). Because these stories also laud the value of sacrifice, retrieval of these stories may have contributed to the many sacrificial answers for the ΔSS variant.

This raises the possibility that the results stem solely from general societal values, and that the analogies were irrelevant. However, against this is the finding that Iranians' choices were strongly connected to whether they found structural commonalities with the cultural base story (in which case they chose sacrifice) or instead found structural differences. Among those who chose the utilitarian choice, significantly more Iranians reported structural differences than reported surface differences from the core story. Thus, even those Iranians who chose the utilitarian option mostly did so by reference to the cultural narrative; they simply considered the structural differences sufficiently serious as to block the analogous inference.

There was no significant difference between ΔST and ΔSV, indicating that a change in sacred values, in this case swapping the roles of the actors, had effects similar to a change in structure. In the Vali story, a person in power helping someone in need by sacrificing his status is considered the moral message of the story.

In conclusion, the results of the first experiment offer some support for the claim that analogical mapping from a cultural moral story to a current dilemma affects participants' decision making when faced with moral dilemmas. The trend of sacrificial decision making among the Iranian participants depended on whether the probe could be structurally aligned to the base moral narrative or not. Because the American control group did not have access to the cultural narrative, structural differences between the variants did not affect their decision making. Of course, it is also possible that Americans may have experienced a different cultural value, that of observing the rules of the game. If so, then they may also have been acting in accord with a set of core values, though not the same set as the Iranians. Future research will investigate this possible difference in core values.

In the second experiment, I examine the effects of a more recent cultural story on people's decision making.

4.5.2. Second Narrative: Hossein Fahmide

For the second study, I used a story about the Iran and Iraq war.

Base story:

During the Iran and Iraq war, Hossein, a young boy who has sneaked into the army, is confronted with a convoy of tanks that, if not stopped, will destroy a part of the city that the boy is fighting at. Hossein can either try to run to his commander on time, inform him about the situation and save his own life or he can stop a tank by sacrificing his own life. Hossein, therefore, took a grenade from a nearby body, pulled the pin out, and jumped underneath the Iraqi tank, killing himself and disabling the tank. This stopped the Iraqi tank division's advance and saved many people's lives.

ΔSF:

During the Bosnian and Serbian war, a young boy sneaks in to the army. One day during the war, he is confronted with a convoy of enemy buses carrying soldiers and weapons. If these buses are not stopped, they will help the enemy destroy part of the city that the boy is fighting at. He can either try to run to his commander on time, inform him about the situation and save his own life or he can stop a bus by running underneath it and activating a mine which otherwise would not work.

ΔST:

During a war, a young boy who has sneaked into the army, is confronted with a tank that if not stopped will destroy a part of the city that the boy is fighting at. He can either try to run to his commander on time and inform him about the attack which would cause the commander to issue a strike from other units against the tanks or he can stop a tank by running underneath it and activating a mine which otherwise would not work.

ΔSS:

During the Bosnia and Serbian war, a young boy sneaks in to the army. One day during the war, he is confronted with a convoy of enemy buses carrying soldiers and weapons. If this bus is not stopped, it will help the enemy destroy part of the city that the boy is fighting at. He can either run to his commander on time, inform him about the situation which would cause the commander to issue a strike from other units against the convoy of buses or he can stop a bus by running underneath it and activating a mine which otherwise would not work.

ΔSV:

During the Bosnian and Serbian war, a young Serbian boy sneaks in to the army. One day during the war, he is confronted with a convoy of Bosnian buses carrying soldiers and weapons. If these buses are not stopped, they will help the Bosnians destroy part of the city that the boy is fighting at. He can either try to run to his commander on time, inform him about the situation and save his own life or he can stop a bus by running underneath it and activating a mine which otherwise would not work.

After reading one of these dilemmas, the participants were asked similar questions to those asked in experiment one, with only the first question being different:

1. What should the young boy do?
    a. Run away
    b. Sacrifice his own life

The control group received translations (with minor changes, e.g. names) of the above variants.

4.5.2.1. Method

The participants and procedure were as in Study 1.

4.5.2.2. Results

The total and conditional retrieval rates are reported in Table 5. Following the results of the first experiment, ΔSF had the highest number of recalls (93%), higher than ΔST (81%) ($\chi^2 = 3.30$, $df = 1$, $p < 0.1$) and ΔSS (77%) ($\chi^2 = 6.70$, $df = 1$, $p < 0.01$). Also, there were higher number of retrievals in ΔSV (90%) than in ΔSS ($\chi^2 = 5.10$, $df = 1$, $p < 0.1$). Therefore, as expected ΔSS had the lowest number of recalls.

Among the control group, 18 participants recalled a story, 12 of which were other stories used in the experiment and the rest were pop-culture movies such as Iron Man and Saving Private Ryan.

Table 5: Retrieval rates of the core narrative for Iranians in the second narrative

Alternative Chosen	ΔSF	ΔST	ΔSS	ΔSV
Utilitarian	0.46	0.52	0.42	0.61
Sacrificial	0.47	0.30	0.35	0.29
Total Retrieval	0.93	0.82	0.77	0.90

The proportion of sacrificial choices to the total number of selected choices for each variant is reported in Table 6. As in the first narrative, Iranians who received the ΔSF variant were highly likely to choose the sacrificial option. For the Iranian group there was a significant difference between the following variants: ΔSF and ΔST ($\chi^2 = 4.2817$, $df = 1$, $p < 0.05$), ΔSF and ΔSV variants ($\chi^2 = 5.6432$, $df = 1$, $p < 0.01$) and ΔSS and ΔSV ($\chi^2 = 4.0652$, $df = 1$, $p < 0.05$). The number of alternate stories retrieved in ΔSS among participants who chose the sacrificial option was marginally more significant than the ΔSF variant (χ^2 = 1.97, *df* = 1, *p* = 0.160). For the control group, there were no significant differences between the different variants[5].

Table 6: Proportions of sacrificial choices to total number of selected choices for the second narrative

	ΔSF	ΔST	ΔSS	ΔSV
Iranian Group	0.50	0.32	0.45	0.30
Control Group	0.33	0.17	0.25	0.22

As in the first narrative, among those who reported structural similarities to the core narrative, a larger number chose the sacrificial choice (62%) than the utilitarian choice (38%) (χ^2 = 7.60, *df* = 1, *p* < 0.01). However, among those reporting structural differences, a larger number chose the utilitarian choice (86%) than the sacrificial choice (14%) (χ^2 = 75.74, *df* = 1, *p* < 0.001). In the other direction, among the Iranians who chose the sacrificial choice, a significantly larger number reported structural similarities to the cultural narrative (52%) than reported surface similarities (21%) (χ^2 = 17.55, *df* = 1, *p* < 0.001). Among Iranians who chose the utilitarian choice, the reverse held: a significantly larger number reported structural differences (54%) rather than surface differences (8%) (χ^2 = 57.81, *df* = 1, *p* < 0.001). Note that even those who chose the utilitarian option still mostly made reference to the cultural narrative. As expected, given that Americans did not know the base story, none of these differences were observed in this group. Among Iranians, a significantly larger number of participants who were reminded of a story chose the sacrificial options (92%), compared to the utilitarian option (83%) (χ^2 = 4.6609, *df* = 1, $p < 0.05$). Comparing the trend of the choices across different variants using logistic regression revealed a significant difference between the two cultures ($z = -2.045$, $p < 0.05$) [6].

4.5.2.3. Discussion

Among the Iranian participants, there was again a clear difference between ΔSF and ΔST/ΔSV variants: participants more often chose the second option in the ΔSF variant than they did in the ΔST or the ΔSV variants. This follows our prediction that people draw inferences suggested by their core narratives when they can be structurally aligned. As in study 1, I found a high rate of sacrificial inferences when *both* structure and surface were changed (ΔSS). I believe that this pattern is again due to the fact that more alternate moral stories were retrieved in the ΔSS condition and they might have contributed to the high number of sacrificial choices.

Also, similar to the results of the first study, modifying the sacred value had the same effect as modifying the structure of the scenario. Altering the role of the sacred value(s) involved in a decision making scenario seems to have structural effects, and this reduces the possibility of analogical inference from the base. There was no difference between ΔSF and ΔSV among the control group, which is expected given that defending Muslim land is not a sacred value for the Americans.

The major difference between the two experiments is the reverse trend in the number of sacrificial answers among the two cultures. The Iranian participants made significantly more sacrificial choices in the first experiment than they did in the second experiment ($\chi^2 = 17.0665$, *df* = 1, $p < 0.001$). However, this was reversed among the American participants, who made more sacrificial choices in the second experiment than they did in the first ($\chi^2 = 23.7252$, $df = 1$, $p < 0.001$). This may be because the American participants were reminded of more stories in the second study compared to the first study.

The findings for the ΔSS variant in the first two narratives differed somewhat from predictions. I suspect that this may have resulted from the large number of close variants of the two narratives that exist in Iranian culture. The third narrative is a cultural narrative which has a single dominant form.

4.5.3. Third Narrative: Dehghan Fadakar

In this experiment, I used another famous Iranian story which is a required reading in the third grade school book of all Iranian children. Therefore, compared to the first story, which is usually told by parents to children and therefore each child might hear a slightly different version of it, and to the second story which the theme of it exists in many Muslim religious stories (such as the story of Ashura), we can assume that all the participants know the same version of this story. In order to pinpoint the differences between ΔSF, ΔST and ΔSS, only three different variants were used.

Base Story:

A farmer is returning home from a day of work carrying an oil lamp. He notices that as the result of a landslide, parts of a railroad just outside of a tunnel has been covered with stones. He walks past the tunnel and realizes that a train is heading towards the tunnel. The farmer has two options, he can either try run to the station on time and inform the station manager and save his own life, or he can put his coat on fire, stand in the way of the train, risk his life and try to signal the train. He chooses the second option and saves the lives of many people.

ΔSF:

A man is going to work carrying a flashlight. He notices that as the result of an earthquake, a bridge has collapsed. He walks past the bridge and realizes that a bus is heading towards the tunnel. He has two options: he can either try to run to the station on time, inform the station manager and save his own life, or he can use his flashlight, stand in the way of the of the bus, risk his life and try to signal the bus.

ΔST:

A farmer is returning home from a day of work carrying an oil lamp. He notices that as the result of a landslide, parts of a railroad just outside of a tunnel has been covered with stones. He walks past the tunnel and realizes that a train is heading towards the tunnel. The farmer has two options, he can either try to run to the station on time and have the station manager reroute the train, or risk his life, by standing on the tracks, which will make him famous in his town and he would potentially receive a cash prize.

ΔSS:

A man is going to work carrying a flashlight. He notices that as the result of an earthquake, a bridge has collapsed. He walks past the bridge and realizes that a bus is heading towards the tunnel. He has two options: he can either try to run to the station on time and have the station manager reroute the train, or he can use his flashlight, stand in the way of the of the bus, risk his life and try to signal the bus, which will make him famous in his town and he would potentially receive a cash prize.

After reading one of these dilemmas, the participants were asked similar questions to those asked in experiment one, with only the first question being different:

1. What should the man do?
    a. Run to the station
    b. Risk his own life

The control group received exact translations of the above scenarios.

4.5.3.1. Method

The participants and materials were the same as in Study 1 and 2.

4.5.3.2. Results

Overall, the results were closely in line with the predictions derived from analogical research. The results for retrieval (question 2) accord with predictions from analogy research in showing the large importance of surface similarity matches for retrieval. The total and conditional retrieval rates are shown in Table 7. Iranians who received variants low in surface similarity to the core narrative—either the ΔSF variant (88%) or the ΔSS variant (86%) showed significantly lower retrieval of the core narrative than did those who received the ΔST situation (98%) (χ^2 = 6.42, df = 1, $p < 0.05$ and χ^2= 9.27, df = 1, $p < 0.01$, respectively).

Among the control group, 16 participants recalled a story: 11 of the 16 were reminded of other stories used in the experiment and the rest were reminded of pop-culture movies.

Table 7: Retrieval rates of the core narrative for Iranians in the third narrative

Alternative Chosen	ΔSF	ΔST	ΔSS
Utilitarian	0.11	0.42	0.46
Sacrificial	0.77	0.56	0.41
Total Retrieval	0.88	0.98	0.87

For inference, as show in Table 8, Iranians who received the ΔSF variant –which preserves the relational structure of the core narrative—were more likely to choose the sacrificial option than those who received the ΔST variant (χ^2= 18.05, df = 1, $p < 0.001$) or the ΔSS variant (χ^2 = 33.25, df = 1, $p < 0.001$). For the control group, there were no significant differences between the different variants[5].

Table 8: Proportions of sacrificial choices to total number of selected choices for the third narrative

	ΔSF	ΔST	ΔSS
Iranian Group	0.83	0.56	0. 47
Control Group	0.46	0.37	0.48

As in Study 1, among participants who noted structural similarities with the core narrative, a larger number chose the sacrificial choice (80%) than the utilitarian choice (20%) ($\chi^2 = 66.02$, $df = 1$, $p < 0.001$). However, if participants reported structural differences, a larger number chose the utilitarian choice (58%) than the sacrificial choice (42%) ($\chi^2 = 6.48$, $df = 1$, $p < 0.01$). In the other direction, among the Iranians who chose the sacrificial option, a significantly larger number reported structural similarities to the base (51%) rather than surface similarities (32%) ($\chi^2 = 10.91$, $df = 1$, $p < 0.001$). Furthermore, among participants who chose the utilitarian option, a significantly larger number reported structural (86%), rather than surface (9%), differences from the base ($\chi^2 = 106.84$, $df = 1$, $p < 0.001$). Logistic regression revealed a significant difference in the trend of answers to these variants between the Iranian participants and the control group ($z = -2.18$, $p < 0.05$)[6].

4.5.3.3. Discussion

In this study, among the Iranian participants there was a clear difference between ΔSF and ΔST/ΔSS variants: participants more often chose the sacrificial option for the ΔSF variant than they did for the ΔST or the ΔSS variants. This follows the prediction that people draw inferences suggested by their core narratives when they can be structurally aligned. For the Iranians, analogical inference from a base moral narrative seems to have been the key process when they were presented with the moral dilemmas.

In the third narrative, retrieval mainly depended on surface similarity, conforming to the patterns found in laboratory studies (Gentner, Rattermann, & Forbus, 1993; Reeves & Weisberg, 1994; Ross, 1989) (and in contrast to the pattern in studies 1 and 2). This difference may be due to the difference in cultural patterns for the two stories. As mentioned above, many different versions of the first two stories (used in different contexts) exist across the culture, whereas for the farmer story we can safely assume that all of our Iranian participants know the very same version of it. Furthermore, the farmer story is mainly used in a single context. Thus, we speculate that this story had fewer near competitors during retrieval from LTM. As a result, surface similarity played a more important role in retrieval than structural relations.

The Americans made more sacrificial choices in the third narrative than they did in the first ($\chi^2 = 20.84$, $df = 1$, $p < 0.001$). This may be because that the rate of recall among American participants was higher in the third experiment than it was in the first (they were often reminded of different movies).

In the next experiment, I use MoralDM to model the above findings about the role of cultural narratives on decision making. Due to the reliance of MoralDM on analogical reasoning, we can capture some specific norms of a culture by adding core cultural stories of that culture to the KB of MoralDM. I will explore this feature of MoralDM in the next experiment, where I use the Iranian cultural stories mentioned above to formally investigate how cultural narratives can affect moral decision making.

4.6. Simulating the Results with MoralDM

The above empirical studies demonstrated that participants presented with a ΔSF variant not only were reminded of the base story, but applied them to the variant scenario, choosing personal sacrifice as the protagonist in the base scenario did. Among Iranians who chose the utilitarian choice a significantly larger number reported structural differences rather than surface differences. Moreover, among the Iranians who chose the sacrificial choice, a larger number reported structural similarities to the cultural narrative than reported surface similarities. American participants, however, not being familiar with the base stories, did not behave differently under any of the variations. These findings followed the main prediction that people draw inferences suggested by their core narratives when they can be structurally aligned.

The key difference between these narratives and scenarios used in the previous chapter is that these stories focus on the intentions of the agent with respect to the choice being made. This may be due to the fact that in Islam intent is “the criterion for value judgments” (Musavi-Lari, 1997). Therefore, many Muslim cultural make explicit that the value of decisions depend on the intentions of the protagonist in the story.

In order to test MoralDM on these stimuli, EA NLU was used to semi-automatically translate the stories and their variants into predicate calculus. However, each of these base stories contained several important additional details compared to the other scenarios: First, similar to the previous MoralDM experiments, each base story included the decision that was chosen by the agent involved in the scenario. Second, the agent’s intention for making that decision was determined and included in the base representation by EA NLU. Third, the moral value involved in the story was also part of the base story. Except for the decision that was made in the story, which is always included in the original Iranian story, EA NLU semi-automatically determined the agent’s intention and the moral value involved in the story (Tomai, 2009a). Intentionality in these stories is focused on the motivation for the chosen action combined with the consideration of the side-effects. For example, consider the Pourya Vali story. It is Pourya Vali’s intentional willingness to risk his status for the purpose of helping the young wrestler and making the old mother happy that is considered a moral exemplar. In the Farmer story, Dehghan Fadakar's intentional willingness to risk his life for the purpose of saving others is the moral value of the story. In the base version of the stories, intention can be inferred from the choice that was made. The main agent in the stories is always presented with a choice and it chooses one of the options, so it can be inferred that he intended the previously mentioned consequences. In the variations of the story, the protagonist's decision is not mentioned in the story. Thus EA NLA infers hypothetical intentionality for the choices. In other words, if Pourya Vali, for example, were to choose the first or second option, then he would be attributed appropriate intentions (Tomai, 2009a). Please refer to Appendix A for the predicate calculus representation of the Pourya Vali scenario produced by EA NLU.

I added the three Iranian base narratives to the KB of MoralDM. Next, for each of the variants presented to MoralDM, both reasoning modules recommended choices. Given that specific rules for these stories were not implemented into the system, the FPR module worked in the calculation-based mode and recommended the choice that would bring the highest benefits to the agent.

The AR module, on the other hand, checked for close analogs. If a similar story was found, candidate inferences were calculated from the base story. A close examination of the stimuli reveals that all the test variants highly resemble their base story. As a result, the similarity score between variants and their base were calculated by the AR module to be high and over the required threshold. However, as stated before, structural differences between a variant and a base story can cause invalid inferences to be mapped to the target. Therefore, if a variant contained structural differences including different protected values or additional structural information not included in the base, then the intention mapped to the target would not be valid or it would not have the same moral value as the one in the base. In the description of the AR module I noted that a previously a case in the KB has to satisfy a number of constraints before it can get accepted as a valid analog. First, the similarity score between the base and the target should be higher than a threshold. Second, they should both have the same order of magnitude relationship between outcome utilities. Third, the inferred moral value of the intention in the target should match that of the base. Please refer to Section 2.4 of Chapter 3 for the details of this reasoning process. Also, the rules involved in reasoning about candidate inferences are listed in Appendix B, Section 2.3. If the moral value of the intention does not match that of the base, then the AR module rejects the inferred candidate inference. As a result, for the variants which have structural differences with the base, the AR module cannot come up with an answer and the answer of the FPR module is accepted. However, in variants which only differed with the base in terms of surface features, the calculated intention of the agent for choosing the sacrificial choice matched that of the base. Therefore, in these cases the system accepted AR module's answer.

To make decisions in ΔSF variants, MoralDM first determines that the variant is a close analog for its base story. Next, the system calculates candidate inferences using the common relational structure of the two representations. The intention for the choice made in the base maps to the hypothetical intention of the protagonist in the target. MoralDM uses candidate inferences to determine the moral value of the intention in the target. In the surface change variants, the inferred moral value of the intention matches that of the base and AR module's answer is accepted. Therefore, MoralDM concludes that the same choice as in the base story, self-sacrifice, is appropriate in ΔSF variants. For example, in the case of ΔSF variant of Dehghan Fadakar, the high moral value of risking one's own life to save some people on a bus matches the moral value of risking one's own life to save people on a train. In the surface variant of Pourya Vali, the inferred moral value of the hypothetical intention of Ali for choosing to lose the match to help the other player get married and make the old mother happy matches the moral value of Pourya Vali in the base story.

Differences in structural change variants deal with side-effects. In the Vali story, the side effect deals with the old mother aiming to use the money to buy expensive clothes, instead of the money making it possible for the other player to buy a house that constitutes this change. In the Farmer story, this change involves willingness to risking one's life for the possibility of fame and fortune, rather than for the sole purpose of saving people. In this variants, MoralDM concludes that the high moral status of, for example, risking one's own life to save other expressed in the base story transfers, by analogy, to risking life for becoming famous and receiving money. This intention is determined by the system to not have the same moral value as the one in the base, and therefore the mapping is invalidated. In variants with structural changes, the result of the FPR module, which is the option that benefits the main agent, is chosen.

All the three stories and their variants exhibit a similar structure of choice and intention, and a similar reasoning process takes place in them. The base stories laud personal sacrifice for a morally valued outcome. In the surface variations, only types are altered. When the base story is retrieved as an analog, MoralDM concludes that the same choice, self-sacrifice, is appropriate in the new scenario. However, the differences in structural changes deal with side-effects and intentions. Therefore, in these variants the mappings are invalidated because the intentions and the moral values do not match and the system chooses the option that maximizes the agent's utility.

MoralDM successfully differentiated between the variants of the three stories, matching the results of the Iranian participants. For each variant the model had to judge whether the protagonist should take the sacrificial choice or the choice that maximized personal utility. In order to model the results of the control group, who were not familiar with the base narratives and as a result did not differentiate between the variants of them, I removed the base narratives from the KB of MoralDM. By doing so, MoralDM operated in the calculation-based mode of decision making and for all variants chose the decision that brought the highest utility to the agent.

4.7. Conclusions

Motivated by the results of Chapter 3, I performed three cross cultural studies to examine whether the processes by which people apply core cultural narratives in their decision making follow the principles of analogical retrieval and mapping. In particular, I investigated how analogical accessibility and alignability influence the use of canonical moral narratives. The results of my experiments suggest that analogical mapping from core cultural narratives can influence moral reasoning about current moral dilemmas. Supporting the hypothesis that analogical processing occurs during moral decision making, my results show some of the keynote phenomena that characterize analogical processing: (1) changes to surface structure of the scenarios did not affect inference from the original cultural stories, once they were retrieved, while changing the deep structure blocked the inference; (2) especially in Study 3, the rate of retrieval of cultural narratives varied based upon the degree of surface and structural similarity with the presented dilemma. One difference from laboratory studies was the very high rate of retrieval overall. I conjecture that this high retrieval rate stems from the importance of the narratives in Iranian culture, as well as from their frequent repetition and schematization (Blanchette and Dunbar, 2000; Gentner et al., in press).

In the last experiment, I used MoralDM to model the results of the psychological experiments discussed in this chapter and by doing so close the modeling loop. In all the variants of the base stories presented to MoralDM, the compositional frame semantics of EA NLU were sufficient to cover the necessary semantic breadth to capture these distinctions (Tomai, 2009a). The combination of analogical reasoning and first-principle's reasoning was proved yet again to be a necessary part of MoralDM. The AR module's result was used in cases where the variant resembled the base story in intention and moral value. However, in cases where these structural features where different from the base, the FPR module took over and chose the option which benefited the main agent by first determining the utility of each option.

In conclusion, the results of this chapter suggest that a core differentiating factor in moral reasoning between cultures may be familiarity with different collections of cultural narratives. Even if the foundations and the logic of morality were universally present, the different cultural stories would cause differences in the judgment of morality between cultures. I believe some well known findings on moral reasoning might be explained by formal examination of moral narratives present within and across cultures.

The preliminary results of this work were published in (Dehghani, Sachdeva, Ekhtiari, Gentner, & Forbus, 2009; Dehghani, Gentner, Forbus, Ekhtiari, & Sachdeva, 2009).

5. Capturing and Categorizing Mental Models using A QP-Based Concept Map System

5.1. Introduction

Qualitative representations capture the intuitive, causal aspects of many human mental models (Forbus and Gentner 1997). This includes aspects of modeling not handled by traditional formalisms, such as conditions of applicability and other types of modeling knowledge. The qualitative reasoning community has explored a wide range of representations and techniques, pursuing its goal to capture the breadth of qualitative reasoning, ranging from the person in the street to the expertise of scientists (Forbus 1996). In previous chapters, I argued for the benefits of qualitative representations in modeling decision making as they provide a useful commonsense approach for comparing utilities with varying degree of sensitivity. The impacts of secular versus sacred values were modeled in MoralDM via qualitative reasoning, using an order of magnitude representation. In this chapter, I examine the use of qualitative representations and analogical generalizations in capturing causal mental models.

As discussed in Chapter 2, causality plays an important role in human decision making. Decision makers develop causal networks and use these networks to connect information and evaluate options (Joyce 1999). Consequently, a formal method for capturing these causal networks is essential for better understanding and modeling the decision making process. Qualitative representations can provide a formal and intuitive method for capturing these causal mental models. Therefore, qualitative modeling could become an important tool for cognitive science, by providing formal languages for expressing human mental models. Formalization provides two benefits: First, we should be able to make predictions about reasoning and decision making strategies based on the structure of individuals' mental models. Second, we should be able to use machine learning techniques to capture common properties of multiple mental models so that we can make generalizations across several individuals within a group or across an individual's mental models at different points in time.

A significant problem with many machine learning techniques is that they require a large amount of data to perform reasonably on categorization problems. This would impose an extreme burden on data collection from human subjects. However, research in cognitive psychology has shown that humans do not require hundreds of examples to perform categorization (Casasola 2005, Gentner and Namy 1999, Nosofsky, et al. 1994).

In this chapter, I present the Qualitative Concept Map system (QCM) which provides a cognitive scientist friendly environment that allows modelers to create and explore qualitative causal models, incorporate them into probabilistic models and output them in formats usable in other forms of reasoning (e.g. analogical reasoning). Next, I examine the use a cognitively motivated model of generalization, SEQL (Kuehne, Forbus, et al. 2000), to perform categorization on QCM models. This method has proven to be especially efficient, requiring an order of magnitude less training data, when it operates on qualitative and highly relational data (Dehghani and Lovett 2006, Lovett, Dehghani and Forbus 2007, Halstead and Forbus 2005, Lockwood, Lovett and Forbus 2008). I examine the use of qualitative representations and analogical generalizations in modeling the similarities and differences in causal reasoning for biological kinds between rural Menominee Native Americans and rural European Americans.

Qualitative Concept Maps are used for modeling and analyzing transcripts of interviews conducted with these groups. These models are then exported as predicated calculus statements and are used to construct generalizations for the groups. These generalizations are tested both by inspection and by creating a classifier to distinguish models from these two cultures. My system efficiently and successfully classified models according to cultural group membership. The results showed that when reasoning about food webs in local ecological systems, rural Menominee adults listed more species and described more interconnections among species than rural European Americans. Furthermore, the system was able to automatically categorize Menominee models based on the level of the expertise of the participants. Overall, the experiments illustrate how qualitative modeling, analogical generalizations and QCM as a modeling tool provide a formal method for capturing causal mental models.

This chapter is organized as follows. First, I describe QCM, discuss its features, and some real-world cognitive science examples modeled in it. Next, I describe the qualitative mode of the system and QRG's qualitative simulator (Gizmo). Then, I describe the probabilistic mode and how information available in the qualitative mode can be integrated into the probabilistic mode. Next, in Section 4.1 of this chapter, I summarize the research on the role of culture in reasoning about biological kinds. Then I use QCM to examine the relationship between culture, expertise, and causal reasoning in the domain of biology. I close by discussing related work.

5.2. Qualitative Concept Map System

QCM is designed as a tool for cognitive scientists. Experimenters can create qualitative representations, explore mental models and integrate these models with other forms of reasoning.

It provides a unified reasoning platform in which mental models can be constructed and analyzed using Qualitative Process (QP) theory (Forbus 1984) and Bayesian Networks (Pearl 2000, 1988). QCM is connected to Gizmo, a full implementation of QP theory, for providing qualitative simulations, including envisionment. Also, QCM uses a Bayesian inference algorithm for calculating probabilities of evidence and posterior probabilities. The user interface of QCM provides easy access to these reasoning features.

QCM uses a concept map interface (Novak and Gowin 1984) and automatically checks for any modeling errors which violate the laws of QP theory and probability theory, providing detailed error messages. QCM can import and export models via GraphML (Brandes, et al. 2001), allowing graphs drawn in QCM to be easily viewed in other graph drawing programs. This facilitates collaboration between modelers. More importantly, for cognitive simulation purposes, models can be exported as predicate calculus statements. This enables QCM models to be used in a variety of types of reasoning, such as analogical reasoning.

QCM has been used for modeling a variety of different phenomena, from abstract models of religious beliefs to concrete qualitative reasoning scenarios. The ideas and features of QCM will be illustrated with models drawn from our modeling work.

5.2.1. QP Modeling

QCM enables experimenters to construct and analyze models using the QP ontology (Forbus 1984). QP theory as a representation language for physical phenomena includes:

- Continuous parameters (*quantities*)
- Causal relationships between them (*influences*)

- Mechanisms underlying physical causality (*physical processes*)

Systems and phenomena are modeled via sets of entities with continuous parameters, whose relationships are expressed using a causal, qualitative mathematics. Processes provide an explicit notion of mechanism. In QP theory direct influences are modeled using I+ (≡ *Increases*) and I- (≡ *Decreases*) which indicate an integral connection between two parameters, i.e., heat flow decreases the heat of its source and increases the heat of its destination. Indirect influences are modeled by $\propto_{Q+}$ (≡ *Influences*) and $\propto_{Q-}$ (≡ *InfluencesOpposite*) which indicate functional dependence between two parameters, i.e., the heat of something determines its temperature.

Using the QP mode of QCM, I modeled the effects of fear on different properties of the self, and effects of external processes on these properties, as described in Jami 'al Sa'adat (The Collector of Felicities) (al-Naraqi 18th century), an Islamic book of ethics written in the 18th century (Figure 14).

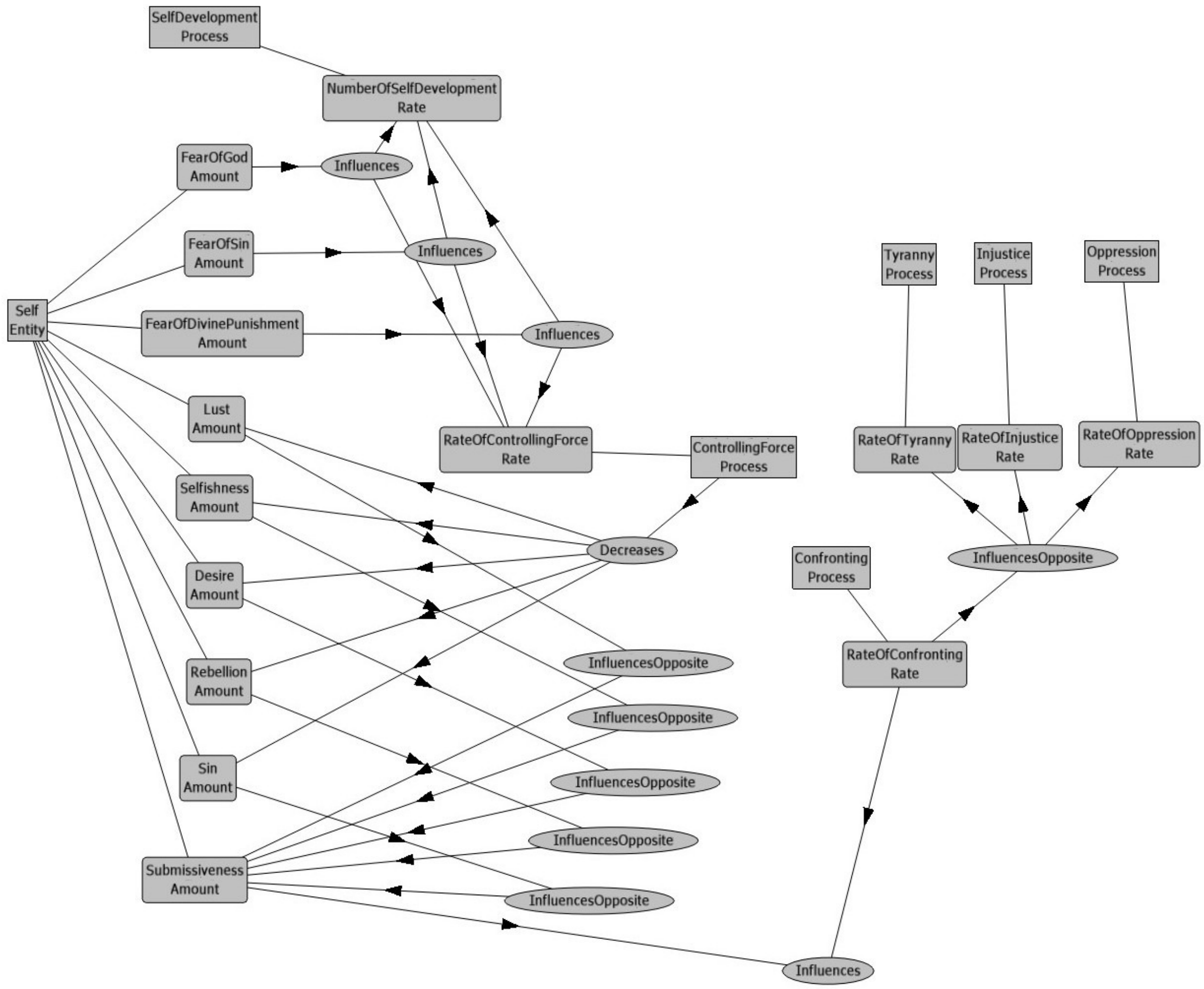


Figure 14: The effects of fear on different properties of the self

Qualitative states capture changes in situations over time modeled by collections of objects and relationships between these objects (Forbus 1984). Often, multiple qualitative states are required to capture change over time. QCM utilizes multiple panes to represent distinct qualitative states. For example, often modelers need to discuss immediate effects of a change followed by long-term effects of a change. The meta-pane (Figure 15) allows modelers to see all the states at once.

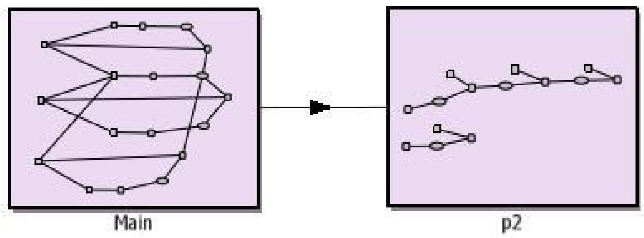


Figure 15: A meta-pane

Modelers can easily extend the vocabulary of specific processes and quantities used in the models, to expedite model creation. QCM is used to build example-specific representations, which is the type of modeling needed for capturing the properties of psychological protocol data.

Figure 16 illustrates one pane from a model for the Bears Disappearing scenario modeled from transcript data, a process which I will discuss in detail in the experiments section. Systems which require general domain theories to be completely specified make the process of modeling difficult and time consuming for novice modelers

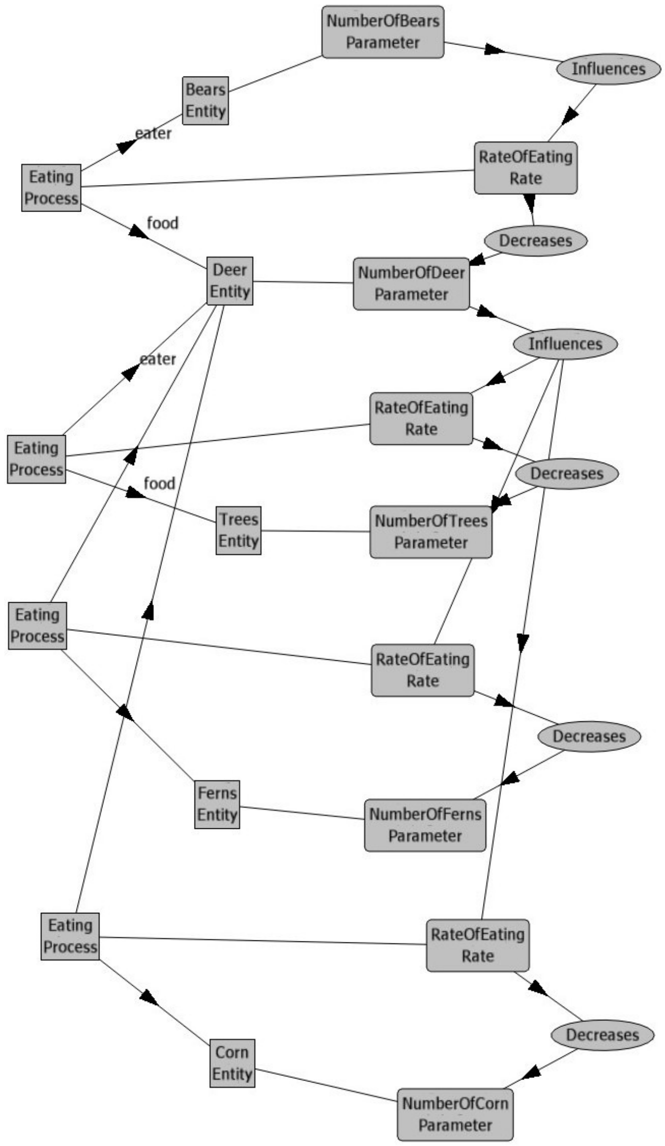


Figure 16: The Bears-disappearing scenario modeled from transcript data

Gizmo Mk2 is a full implementation of QP theory and works as the qualitative reasoning engine of QCM. Gizmo has been designed to be lightweight and incremental. The user has tight

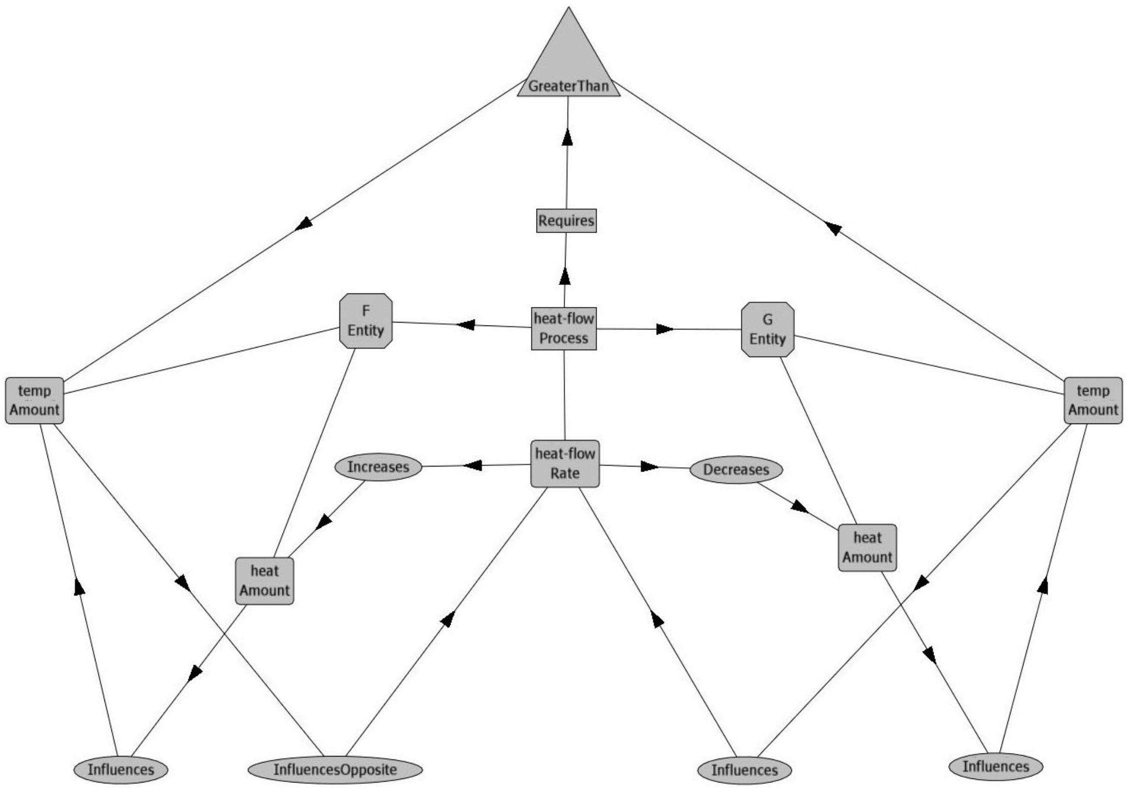


Figure 17: Heat-transfer scenario

control over the process of qualitative simulation in Gizmo. Algorithms for both total and attainable envisioning are included as well. The domain theory and the scenario of the model are automatically extracted from the graph and sent to Gizmo. This extraction is performed by going over all the nodes in the graph and, for each node, determining the type of node it is (e.g. *Entity*, *Process*, *Quantity*). Based on this information, QCM automatically obtains the required information for that type of node from the graph and sends the information to Gizmo. Figure 17 shows the initial state of a heat transfer scenario. The domain theory extracted for the heat-transfer model of Figure 17 is presented in Figure 18. If the system determines that the model is missing some required information, a detailed error message is presented to the modeler.

```
;;; Quantity Functions
(defquantityfunction Rate (?thing))
(defquantityfunction heat-flow-rate (?Rate))
(defquantityfunction heat (?Amount))
(defquantityfunction Amount (?thing))
(defquantityfunction temp (?Amount))

;;; Entities
(defentity G-type
  :quantities   ((heat :type Amount)
                 (temp :type Amount))
  :consequences ((qprop (temp G-type)
                        (heat G-type)))
  :documentation "finite-thermal-physob")

(defentity F-type
  :quantities   ((heat :type Amount)
                 (temp :type Amount))
  :consequences ((qprop (temp F-type)
                        (heat F-type)))
  :documentation "finite-thermal-physob")

;;; Processes
(defprocess heat-flow
 :participants ((the-G :type G-type)
                (the-F :type F-type))
 :conditions   ((> (temp the-G) (temp the-F)))
 :quantities   ((heat-flow-rate :type Rate))
 :consequences ((i- (heat the-G) heat-flow-rate)
                (i+ (heat the-F) heat-flow-rate)
                (qprop (heat-flow-rate heat-flow)
                       (temp the-G))
                (qprop- (heat-flow-rate heat-flow)
                        (temp the-F))))
```

Figure 18: The domain theory generated from the heat-transfer scenario

The automatic extraction of the domain theory and the scenario file is, I believe, a major boon to novice modelers. While many of the ideas of qualitative modeling come natural to scientists, outside of computer science, experience in writing logically quantified formulae is rare. Modelers need motivation, and being able to get results without having to first write a general domain theory helps reduce the entry barrier. As their models become more complex, the automatically produced models can become a starting point for writing standard QP theory domain models.

5.2.2. Bayesian Modeling

Agents continually update their beliefs using different types of new information. These updates affect their causal beliefs about the uncertainties in the world. In order to model this process, we need a rich causal representation and a method for capturing and updating uncertain beliefs about the world. QP theory provides us with a high level of expressiveness needed to capture many intuitive, causal aspects of human cognition. One can use the QP framework to reason about relations between things and the effect of these relations on the state of the world. However, QP theory does not provide any mechanism for capturing probabilities. Bayesian networks (Pearl 2000, 1988) are the most widely used approach for probabilistic reasoning. This formalism provides a succinct representation for probabilities, where conditional probabilities can be represented and reasoned with in an efficient manner. Providing an interface in which both QP and Bayesian formalisms can be used in parallel can potentially be helpful for cognitive scientists.

QCM enables the agent's knowledge about the causal structure of the world to be captured using QP theory, while the agent's uncertain knowledge and expectations about the outcomes of its actions are captured by subjective probabilities, represented by a Bayesian Network. Modelers can switch the mode of reasoning from QP to Bayesian and make probabilistic models. This feature allows cognitive scientists to take advantage of both formalisms. In the Bayesian mode, modelers can perform exact inference on the network and calculate the probabilities using Recursive Conditioning (RC) (Darwiche 2001). RC is an any-space algorithm which works by recursively partitioning the network into smaller networks using conditioning and solving each subnetwork as an independent problem. Networks created in the Bayesian mode are saved in the

Hugin format which is the standard format for many data mining and machine learning programs. This again helps modelers who use QCM collaborate more easily with other scientists using other modeling programs.

5.2.2.1. Determining a Priori Probabilities using Qualitative Simulations

One of the main obstacles in probabilistic reasoning is finding the a priori probabilities of variables in the model. One approach to overcoming this obstacle is to use qualitative simulations. QCM uses the information available in the QP mode to calculate a priori probabilities of quantities used in the qualitative model. In this framework, the probability distribution is defined over a set of possible worlds determined by the constraints of the qualitative model. This approach assumes uniform distribution over all qualitative states. If the modeler chooses to include a qualitative parameter, such as a quantity or a derivative of a quantity available in the qualitative model, as a node in the probabilistic model, QCM can determine the probabilistic distribution of the values of that parameter by model counting. That is, by calculating the degree of belief in that statement over all the possible worlds determined by qualitative envisionment. For example, if `(temp F) > (temp G)` relationship from the heat-transfer scenario of Figure 17 needs to be included as a node in the model, QCM performs an attainable envisionment determining in how many possible worlds `(temp F)` α `(temp G)` where α = {<, <=, =, >=, >, ?} hold to be true. Based on this measure a probability value can be assigned to `(temp F) > (temp G)` (see Figure 19 for an example of a Bayesian network which uses this relationship). In other words, we are saying that under the current constraints in *n* of *m* possible worlds `(temp F) > (temp G)`, therefore the probability of `(temp F) > (temp`

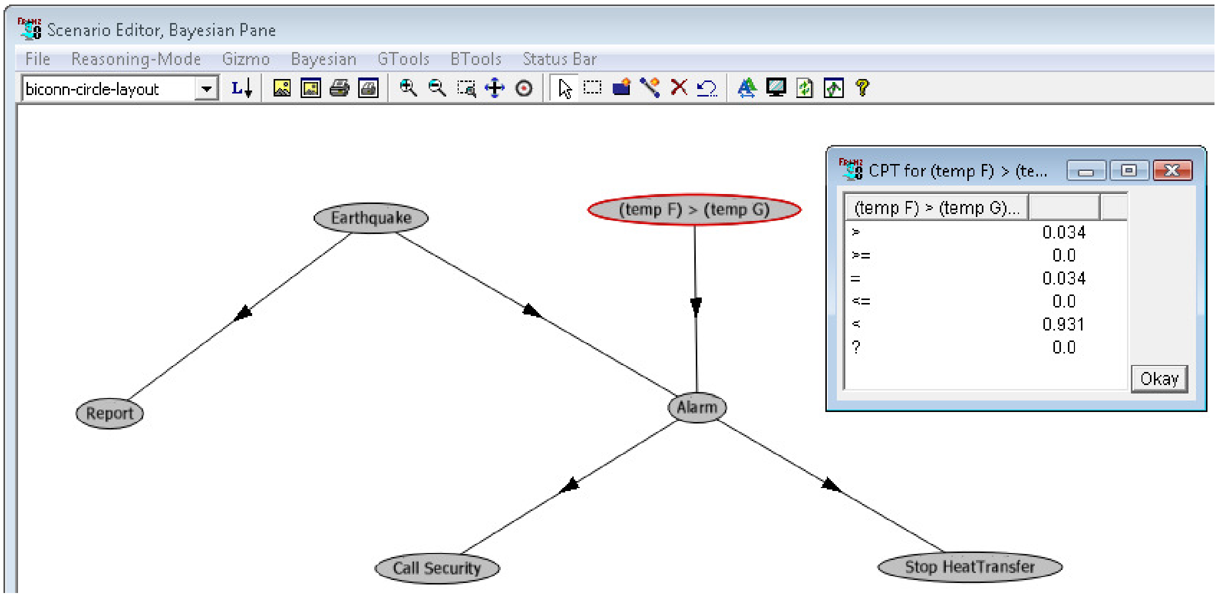


Figure 19: A Bayesian network

G) is *n*/*m*. I believe this method can provide a robust way of estimating a priori probabilities for physical phenomena for which we can define a QP model.

5.3. Automatic Classification of Models

Here I describe a method for building generalizations from QCM models. These generalizations make explicit the common structure found in the models. They can also be used to automatically categorize subsequent QCM models.

As discussed in the introduction, traditional machine learning techniques require a large amount of data to perform reasonably on classification problems. This places an unreasonable burden on cognitive scientists who are seeking to use them for data analysis. However, there is clear psychological evidence that people do not require hundreds or even tens of examples to learn many categories. This phenomenon repeats across domains from spatial language learning (Casasola 2005) to image classification (Gentner and Namy 1999, Nosofsky, et al. 1994).

Gentner and Loewenstein (2002) have argued that individuals learn categories through a process of progressive abstraction, wherein instances of a category are compared and the commonalities are abstracted out as a direct result of the comparison. In many cases, the commonalities resulting from comparison appear to be in the relational structure of the cases being compared. In this section, I use a cognitively plausible supervised learning method based on generalization from relational case descriptions. This method is domain-general and typically requires an order of magnitude fewer examples than other learning algorithms. The major benefit of this technique is that, although it only requires very small training sets, utilizing highly relational structures it can achieve the performance of machine learning algorithms which require orders of magnitude larger data sets.

5.3.1. Comparison and Generalization

SEQL (Kuehne, Forbus, et al. 2000) is a model of generalization built on SME (see Chapter 3). SEQL is based on the idea that when humans are exposed to multiple exemplars of a category, they construct a generalization by comparing the exemplars and abstracting out the common structure. SEQL does this by comparing individual cases using SME. For each category, SEQL maintains a list of generalizations and exemplars. Each new incoming exemplar is compared against the existing generalizations, and if it is sufficiently similar, the generalization is refined based on their common structure. Otherwise, the exemplar is compared against other, unassimilated exemplars. If it is sufficiently close to one of them, a new generalization is formed from their common structure. Originally non-overlapping structure was simply thrown away. Now, SEQL associates a probability with every expression in a generalization which is updated with each new exemplar, and it only gets rid of very low-probability structure (Halstead and Forbus 2005).

SEQL is capable of performing both supervised and unsupervised learning. For unsupervised learning, the user can simply present it with a set of cases, and it will group similar cases together and form generalizations from them. One method for performing supervised learning using SEQL is to set the assimilation threshold to be extremely low, essentially telling SEQL that all the cases being considered belong in the same generalization. SEQL will then construct a single generalization based on what is common among all the cases being considered.

SEQL has previously been used in automatic sketch recognition (Lovett, Dehghani and Forbus 2007), automatic music genre classification (Dehghani and Lovett 2006), classifying terrorist activities by perpetrator (Halstead and Forbus 2007), learning spatial language (Lockwood, Lovett and Forbus 2008) and modeling conceptual change (Friedman, Taylor and Forbus 2009, Friedman and Forbus 2008). The major benefit of SEQL is that, unlike most machine learning algorithms, it only requires very small training sets. Given the cost of running human participants and analyzing data (even with QCM simplifying the process of formal model construction), extracting as much as feasible from available data is important.

5.4. Experiments

The purpose of the following experiments is to further examine the relationship between culture and expertise in ecological reasoning, and to use emerging differences as a basis for testing the effectiveness of QCM modelling. Culture is defined here as the causally distributed patterns of mental representations, their public expressions, and the resultant behaviours in given ecological contexts (Atran, Medin and Ross 2005, Sperber 1996, 1985). People's mental representations interact with other people's mental representations to the extent that those representations can be physically transmitted in a public medium (language, dance, signs, artifacts, etc.). These public representations, in turn, are sequenced and channelled by ecological features of the environment (including the social environment) that constrain interactions between individuals.

The cultural communities involved in the present work include rural Menominee Native Americans and rural European Americans. The Menominee live on 234,000 acres of heavily forested land along the Wolf River in Northeast Wisconsin. The European Americans involved in this research live in the neighboring town of Shawano. Although members of the two communities engage in similar outdoor activities, such as hunting, fishing, and berry-picking, Menominee individuals are more likely to engage in culture-specific ceremonial practices outdoors and are also more likely to simply engage in 'observing' practices (e.g., walks in the forest, whereas rural European Americans are more likely to engage in outdoor sporting activities (e.g., fishing competitions) and outdoor work-related activities (e.g., landscaping) (Bang, Medin and Atran 2007).

The following experiments examine the similarities and differences in causal reasoning for biological kinds between and within these two cultures. We can get a more objective perspective on these differences using automatically constructed generalizations of field data. By examining analogical generalizations created from mental models of people from particular groups, I show that we can concisely summarize common properties of the models of people and cultures.

Moreover, the results of my experiments indicate that analogical generalization can provide a valuable analyses tool for social science research.

Before turning to descriptions of the experiments, analyses, and results, I summarize some research on the role of culture in reasoning about nature and biological kinds, and describe the psychological experiment which provided the data for the QCM experiment.

5.4.1. The Role of Culture and Expertise in Reasoning about Biological Kinds

There are many reasons to believe that there might be similarities in individuals' causal understanding of relationships in nature. Medin, Atran, and their colleagues (Atran, Medin and Ross 2005, Medin and Atran 2004), building on decades of important work in ethnobiology, have found that, in spite of highly varying input, a few key principles guide the recognition and organization of biological information in extraordinarily similar ways. For instance, there is marked cross-cultural agreement on the hierarchical classification of living things, such that plants and animals are grouped according to a ranked taxonomy with mutually exclusive groupings of entities at each level (Atran 1990, Berlin, Breedlove and Raven 1974, Berlin, Breedlove and Raven 1973, C. Brown 1984, Hays 1983, Hunn 1977). The highest level of taxonomic organization includes the most general categories, such as the folk kingdom rank (which includes groupings such as plants and animals), and lower levels distinguish between increasingly greater degrees of specificity (e.g., life forms such as tree or bird; generic species level such as oak or blue jay). There is cross-cultural agreement that the appearance and behavior of every species is caused by an internal biological (and usually unspecified) essence that is inherited from the birth parents and is responsible for identity persistence in the face of physical and developmental transformation (Atran 1998, Atran, Estin, et al. 1997, S. A. Gelman 2003, Gelman and Wellman 1991, Medin and Atran 2004, Sousa, Atran and Medin 2002).

However, there is also evidence suggesting considerable variability within these universal constraints in folk biological concept formation as a function of both experience with the natural world and cultural salience (two highly related factors). For instance, Rosch and Mervis (1975) have found that the life form level is the level for which urban undergraduates possess the greatest knowledge (i.e., basic level), but Berlin (1992) found that among traditional societies in which individuals have more direct experience with the natural environment, the basic level corresponds to the generic-species level, and these differences have been attributed to differences in expertise (Medin and Atran 2004). Other findings implicate cultural differences above and beyond expertise. For instance, Menominee Native Americans are more likely than rural European Americans to see themselves as a part of nature rather than apart from nature and to say that every creature has a role to play on Mother Earth (Bang, Townsend, et al. 2005).

When asked to sort biological kinds into categories, individuals from different communities vary not only in their taxonomic sorting but also in the degree to which they spontaneously sort along ecological dimensions, and this difference is not as predictable on the basis of expertise or experience alone. Specifically, Medin, Ross, Atran, Burnett, and Blok (2002) found that Menominee Native American fisherman and European American fishermen, who both have similar levels of expertise about fish and fish habitats, exhibit differences in ecological sorting of fish during a regular sorting task. Menominee fishermen are significantly more likely to sort in terms of ecological relationships (e.g., *river fish*, *bottom feeders*). In contrast, European American fishermen were more likely than Menominee fishermen to sort fish on the basis of morphological or taxonomic information (e.g., *bass family*). This pattern was found for both expert fishermen and for nonexperts in the two communities. Furthermore, in a subsequent task involving questions about fish-fish interactions, Menominee fishermen were significantly more likely to report positive and reciprocal relations, although both groups were equally likely to report negative relations.

Similar differences in ecological reasoning were found for children from these communities, such that Menominee children were more likely to reason about shared properties between living things on the basis of ecological relations, relative to rural European American children (e.g., a bee and a bear share an internal property because a bee makes honey and a bear eats honey) (Ross, et al. 2003). These results suggest that direct experience with the natural world influences knowledge and salience of ecological relations. In general, a substantial amount of research across cultural groups cannot be limited to explanations involving either expertise or other cultural factors, but instead a confluence of experience-based and culture-based factors in folkbiological thought.

Although prior research suggests that there are cross-cultural differences in causal models, little research has focused on directly assessing such differences. We used qualitative modeling and analogical generalizations to investigate the relationship between culture and expertise in ecological reasoning**.** Our collaborators (Doug Medin's Group) interviewed experts (i.e., hunters and fishermen) and novices (individuals who do not hunt or fish) from Menominee Native American and from European American cultural communities. Participants were presented a scenario in nature and were asked open-ended questions about potential consequences of perturbations to species' populations within an ecological system (see Appendix E for an interview script). Transcriptions of three scenarios were modeled in the present study. In each scenario, participants were told about a perturbation in an ecological system and were asked to

Do you think that the disappearance in the bears would affect other plants and animals in the forest?

-Probably just like shrubs and stuff that these animals the basic food sources like berry plants and stuff. And then maybe larger trees, too, because bears climb trees.

…

Because of there's a more competing for water in the soils. There's more shade, because I'm assuming it's a taller tree. So there's more shade so the ground growth couldn't grow as well. It would provide more nesting areas for the animals that use it for nesting. So they might benefit from it but they'd have less food.

-Right. And so do you think that other trees would be affected?

Figure 20: Excerpts from a transcript

speculate about the effects of such an event on other plants and animals in the forest. In one scenario, the perturbation involved the disappearance of all of the bears in a nearby forest. In another scenario, the perturbation involved a doubling of the bear population in a nearby forest. In a third scenario, the perturbation involved the disappearance of all of the poplar trees in a nearby forest. Each participant was presented with all three scenarios. After each scenario, participants were first allowed to openly discuss any consequences that came to mind before being probed with an exemplar (e.g., eagle) that represented a particular trophic type with respect to the perturbation species (e.g., competitor). Given the open-ended nature of the interviews, the number of probes presented to participants varied across individuals depending on the depth of initial responses and the degree to which they responded to subsequent probes.

The verbal explanations of the subjects were transcribed (see Figure 20 for example), and used as data to construct formal qualitative models expressing their beliefs (see Figure 16 for example). I then examined whether analogical generalizations could accurately classify individuals according to culture and level of expertise. Using analogy to automatically construct generalizations from field data should provide a highly systematic analysis of differences in causal models. Two experiments were conducted to test the hypotheses that the generalizations constructed from formal models of food webs can be used to classify participants according to culture and level of expertise. In both experiments, QCM was used to model 81 transcripts, generated in response to three food web scenarios. I then used SME and SEQL (described above) to automatically classify the data. The data were also coded manually and examined to determine the extent to which we could find overlap between manual and automatic analysis techniques.

Based on previous research cited above, we predicted that Menominee participants would list more specifies as being affected by the perturbation in the ecological system, and would list more interconnections among species affected by the perturbation, relative to European American participants. In other words, Menominee causal models should be more inclusive and interconnected than European American causal models. Furthermore, within each cultural group, we expect to see a difference between causal models of experts (hunters) and non-experts, such that experts' models are more inclusive and interconnected.

5.4.2. Constructing Formal Models from Transcripts

I transcribed the verbal responses to the interview questions described previously and used these transcripts to make QCM models. I used the QP vocabulary for modeling the causal networks described by the participants. First, I looked at what the participants reported were the agents of cause, or what initiated the causal changes in the network. These agents of causal changes were modeled as processes. Next, I identified the entities involved in the causal network. That is, I looked at what 'things' went through changes in a described scenario. Both entities and processes have parameters (Rate, Amount, and Level) associated with them. I then recognized these parameters and modeled them in the network. In my framework, edges connecting processes to other variables are considered direct causal relations. Next, I proceeded by identifying direct causal relations, increase and decrease, between the processes and the entities described in the transcripts. After that, I modeled indirect causal relations. These relations include influence and infuenceOpposite and indicate causal changed between two entities. I then verified that the constructed model followed the verbal response of the subject.

5.4.3. Experiment 1

Can the generalizations produced by SEQL be used as models of each culture, to classify data from new individuals as to what culture they belong to? To answer this question, I conducted a series of trial runs in which the models were randomly divided into training and test sets. On average 6 models from each group were used for the training set and 2 models from each group for the test set. In each run, I used SEQL in supervised learning mode to produce two generalizations from the training set, one for Menominee participants and one for European American participants. These generalizations were then used to classify models in the test set by using SME to compare each model with the two generalizations. I calculated the percentage of the model's expressions that aligned with each generalization, and the percentage of the generalization's expressions that aligned with the model, and classified test models based on which generalization it had more in common with. Figure 21 shows the overall structure of the system. *M* and *A* respectively represents models built from Menominee transcripts and European American transcripts. *X* represents a random model from the test set. I tabulated successful classification by cultural group and averaged the results over all trials. To reduce the chance of

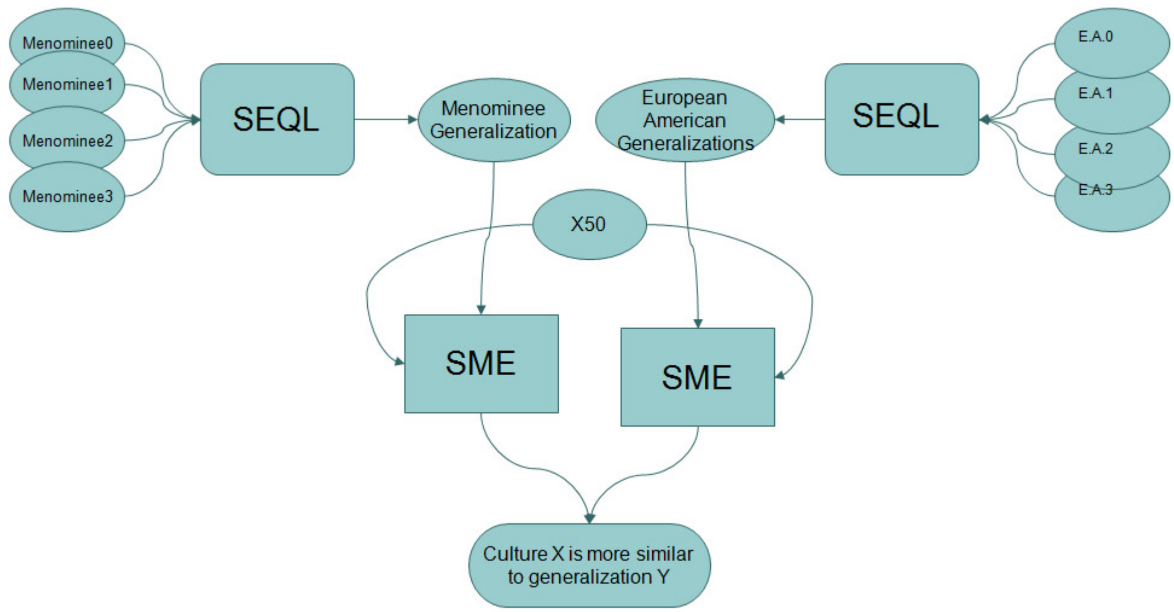


Figure 21: Overall structure of the system

random bias, I conducted 1,000 trial runs (i.e., random training and test sets were generated 1,000 times) and averaged the results.

5.4.3.1. Results

Table 9 shows the results of the first experiment. In the first column the percentage of Menominee models correctly classified as Menominee is shown. In the second column, the percentage of European American models correctly classified as European American is shown. The last column shows the overall accuracy of the system. The average accuracy across the three scenarios was 64%, where a random classification would result in 50% accuracy and 55% would be considered highly significant ($p < 0.001$).

Table 9: Performance of the classification system for the first experiment

	Menominee	European American	Overall Accuracy
Bears Disappearing	65%	57%	61%
Bears Doubling	82%	52%	67%
Poplar Disappearing	64%	64%	64%

5.4.3.2. Discussion

My system was able to automatically compute generalizations which differentiated between the two cultures. It was also able to find similarities in causal models from the same culture. By examining the system's results, we can gain insights into the differences and similarities between the models. Specifically, I found that the number of entities that were consistent across individuals was higher in Menominee models. I examined the generalizations from a single test run for each scenario, in which the system achieved 70% accuracy. For this test run, there were 24 entities found consistently across all Menominee models vs. 16 entities for American European. Also, the number of consistent causal relations was higher among Menominee. Menominee models contained 4 causal relations found consistently across all models, whereas European American models only contained 2.

As per our prediction, the generalizations that were made from Menominee models were more detailed, larger and therefore subsumed other smaller generalizations. This had the unfortunate side-effect of biasing models towards being classified as Menominee. This effect was most salient in the Bears Doubling scenario. However, as mentioned above, the open-ended nature of the interviews led to variation in the number of questions presented to participants across individuals, and the resultant variability in responses can introduce some difficulty when attempting to evaluate similarities in causal maps.

5.4.4. Experiment 2

How are the causal models of hunters (experts) different from non-hunters (novices)? Can we use these differences to automatically classify the models from new individuals as to level of expertise? The overall structure of the system was the same as Experiment 1. The models from each culture were divided into two groups. The first group included models from hunters and the second group included models from non-hunters. In this experiment, I tabulated successful classification by expertise level and averaged the results over all 1,000 trials. For each culture, in each run, 5 models from experts and 5 models from non-experts were used to make the classifiers and 2 models from each group were used to test these classifiers.

5.4.4.1. Results

My system was able to correctly classify experts from non-experts within Menomonee models 72.5% of the time (55% would be considered highly significant, $p < 0.001$). Models from experts in general were more inclusive and had more depth and breadth to them. On the other hand, my system failed to classify experts from non-experts within European American models, with an accuracy of 52% ($p = 0.22$). This result suggests that although there were variations within the

European American models, these variations were not due to differences in the level of the expertise.

5.4.4.2. Discussion

Manual quantitative analyses of the data after the computational experiment revealed that when the number of relations were counted for each individual participant and averaged within groups, Menominee hunters were significantly more likely than Menominee non-hunters to mention ecological relations (19.80 vs. 10.14, respectively; $p < .01$). In contrast, within the European American sample, there was no difference in the number of ecological relations mentioned by hunters and non-hunters (16.08 vs. 16.22, respectively; $p = .97$). Using our method, we were able to categorize and discover these differences in a more systematic and faster fashion.

5.5. Related Work

QCM is a successor to VModel (Forbus, Carney and Harris, et al. 2001, Forbus, Carney and Sherin, et al. 2004). VModel was developed to help middle-school students learn science. Like QCM, it uses a subset of QP theory to provide strong semantics. However, VModel was limited to single-state reasoning, whereas QCM can be used to model continuous causal phenomena with multiple states. Similar differences hold with Betty’s Brain (Biswas, et al. 2001), which provides a domain-specific concept map environment that students can use in learning stream ecology.

The closest other qualitative modeling tools are MOBUM (Machado and Bredeweg 2001) and VISIGARP (Bouwer and Bredeweg 2001) which have lead to Garp3 (Bredeweg, Bouwer, et al. 2007, Bredeweg, Salles, et al. 2006). Like QCM, these environments are aimed at researchers, but their focus is on constructing models for qualitative simulation, including generic, first-principles domain theories. QCM focuses instead on helping capture concrete, situation-specific qualitative explanations of phenomena. Thus, it provides a useful tool for scientists working with interview data.

Different approaches for qualitative Bayesian inference have been proposed. These methods include: qualitative probabilistic networks (Wellman 1990), qualitative certainty networks (Parsons and Mamdani 1993) and a method which incorporates an order of magnitude reasoning in qualitative probabilistic networks (Parsons 1995). Keppens (2007a, 2007b) employs some of these methods for qualitative Bayesian evidential reasoning in the domain of crime investigation. QCM integrates information available from qualitative simulations in probabilistic networks, whereas other approaches mostly use qualitative techniques in performing inference on Bayesian networks.

Recently, computational modeling of cultural attributes has gained increasing attention. I discussed one of these models, CARA (Subrahmanian, et al. 2007), in Chapter 3. Dickerson, Martinez, Reforgiato and Subrahmanian (2008) propose a multi-player online gaming architecture which simulates different characteristics of certain cultures. This framework has been developed to help members of the US military "understand how best to reason about a particular part of the world" by playing different "culture games". This architecture was built by researchers working on the CARA system; therefore it shares many commonalities with that system, including a stochastic opponent modeling module and a real-time opinion extractor module. Martinez, Simari, Sliva and Subrahmanian (2008) use two algorithms based on vector similarity to predict a group's behaviors based on the data on the groups' past behaviors. These classes of algorithms, called CONVEX, calculate the similarity between two situations using different mathematical distance functions, including Euclidean and Canberra distances, of vertices representing two situations. These vertices carry different numerical information about characteristics of a group's behavior in a certain situation. The authors do not mention how these numerical values are compiled and calculated. Shutters and Cutts (2008) simulate the effects of introducing arbitrary cultural traits, such as moral values, in social network structures. Their agent-based model is used to explore various different parameters, such as time and network structure, for which a society may reach consensus with respect to that trait. Miller, Wu, Funk and Johnson (2007) propose a computational model of etiquette and politeness in a culture. Their model is based on Brown and Levinson's (1987) anthropological cross-cultural studies of politeness. This model predicts the level of politeness that a hearer will expect based on three parameters: relative power of the hearer over the speaker, social distance between the two agents and imposition of the act. These approaches, although important steps towards computational modeling of cultural behavior, are orthogonal to the research presented in this chapter. My approach consisted of using a qualitative modeling tool to model psychological interview data and applying analogical generalization to find common cultural traits within these models.

Traditional machine learning techniques require very large datasets to perform well on classification problems, and classification from small number of data points is still an open problem in machine learning. There have been a few attempts to solve this problem. For example, Plumbley (1994) proposes a neural network approach for learning from small data sets. This approach is computationally expensive, and to my knowledge has not been implemented.

Tengli, Dubrawski, and Chen (2005) present a method for handing outliers in small datasets. Their method is used to learn propositional rule lists, if-then-else lists, and is unable to handle higher order rules.

Much of machine learning literature in the last 20 years has focused on learning from tabular data and feature vectors. Only recently has concerted effort been made to move away from this paradigm. Dayanik and Nevill-Manning (2004) apply graph partitioning techniques to cluster relational data and perform learning. Getoor et al. (2001) propose an approach based on Bayesian reasoning which learns probabilistic features of objects and links between these objects. Blockeel and Uwents (2004) propose a neural network approach for learning relations between the data. By contrast, the approach applied in this chapter uses independently tested cognitive models of analogical matching to build such a unifying relational schema, the generalization, from the examples provided. This approach allows learning relations higher than first-order and it requires orders of magnitude fewer examples.

5.6. Conclusions

QCM provides the basic functionality needed for cognitive scientists to build, simulate and explore qualitative mental models. It offers a friendly interface for experimenters to explore causal models using QP theory semantics (Forbus 1984). QCM uses Gizmo as its qualitative reasoning engine, offering a full range of qualitative simulation abilities. Modelers can also work in a probabilistic mode and use RC to perform exact inference on their models. QCM automatically integrates qualitative information for calculating a priori probabilities of quantities used in the qualitative mode. The interface of the system has been enhanced offering easy access to reasoning capabilities. Models can be exported in different formats facilitating collaboration between modelers. I believe that QCM provides the formalism and the functionality necessary for automatic evaluation of psychological data about causal mental models. Moreover, it can potentially be a helpful tool for teaching undergraduate cognitive science courses.

In previous chapters, I showed how qualitative reasoning can be an intuitive method for comparing utilities of choices. In this chapter, I used QCM to create formal models based on transcripts of the interviews and showed that cultural differences in causal reasoning about food webs can be captured to some degree in terms of similarities and differences in qualitative models extracted from transcript data. Qualitative representations proved to be an intuitive and valuable formalism for capturing these causal mental models. QCM as a modeling tool ultimately helped illuminate our understanding of the relationship between culture and expertise in a formal and concise fashion.

Although previous manual analysis of the transcripts have shown to be very difficult and time consuming, by using SEQL and SME I was able to find similarities and differences by automatically constructing generalizations of causal models built from the transcripts. I employed a classification method based on human analogical reasoning which can be trained with data sets that are orders of magnitude smaller than the current requirements of general machine learning algorithms. The system can efficiently learn probabilistic generalizations from relational descriptions and use these generalizations to classify QCM models. The results of my experiments proved that analogical generalization can be a valuable tool for analyzing data from social science research.

In conclusion, in this chapter I have argued that qualitative reasoning combined with analogical generalization is a promising approach for capturing and classifying causal mental models. Moreover, QCM as a modeling tool provides a cognitive scientist friendly environment for exploring these qualitative causal models.

The preliminary results of this work were published in (Dehghani & Forbus, 2009; Dehghani, Unsworth, Lovett, & Forbus, 2007).

6. Closing Thoughts

In this dissertation, I argued for the importance of highly structural representations in conjunction with analogical and causal reasoning for capturing and modeling the effects of culture on cognition. Specifically, focusing on moral decision making, I described MoralDM, which relies extensively on structural representations and intergrades analogical, qualitative and first principle reasoning techniques. Moreover, I demonstrated the role of analogical reasoning and cultural narratives in moral decision making. The last part of this thesis examined the role of culture on causal reasoning about biological kinds using a modeling tool called QCM in combination with computational models of analogy and categorization. In this final chapter, I briefly review the major claims of this thesis, discuss future work and close with some final thoughts.

6.1. Integrated Model of Moral Decision Making

As discussed throughout this dissertation, recent psychological findings have shed light on the process of human decision making by showing predictable violations of axioms of economic theory. Moreover, these results indicate that a single process or a single mode of decision making cannot capture the full spectrum of human decision making (Bennis, Medin, & Bartels, in press; Hastie, 2001). However, the majority of models of decision making in AI operate solely on utility economics and overlook many of these findings. For instance, none of these models reflect on past experiences, use background knowledge or model the effects of culture on decision making. Moreover, the majority of these models operate on propositional logic and lack the level of expressiveness needed to model human decision making. Many of them apply computationally expensive reasoning methods or require very large search spaces which result in the system becoming intractable when dealing with real world decision making problems.

In Chapter 2, I described MoralDM, which is the first cognitively motivated integrated model of recognition-based moral decision making. MoralDM integrates several AI reasoning methods and models psychological findings about cost-benefit analysis and deontological types of reasoning. It takes as input a scenario in natural language and chooses a decision which is morally preferred for a given culture. In order to reduce tailorability, MoralDM uses EA NLU (Tomai, 2009a; Kuehne, 2004) for producing formal representations from the majority of the test stimuli. It employs both first-principles reasoning and analogical reasoning to model known findings on moral decision making. If a solved scenario resembles the target case in structure, using analogical inference, MoralDM imports knowledge from the base scenario to the target and uses the imported knowledge to choose a decision for the target case. The varying degree of quantity sensitivity toward outcome utilities is modeled via qualitative reasoning, using an order of magnitude representation. I tested this model on stimuli from several psychology experiments. In conclusion, I argued that moving away from utility based models and applying integrated techniques help us both study the underlying processes of human decision making, and give our models the ability to tackle a broader range of problems. As a result, an integrated reasoning approach can help solve some of the shortcomings of existing models of decision making in AI.

6.1.1. Future Work

I plan to pursue several lines of investigation next. First, I plan to test MoralDM on a wider range of moral dilemmas, using data gathered from participants from multiple cultural groups. This will require extending the first-principles reasoning rules to cover a broader range of scenarios. Constructing these rules is a time consuming and error prone process. One alternative is to automatically extract rules by generalizing over previously made decisions. This can be done by running SEQL (Kuehne, Forbus, Gentner, & Quinn, 2000), a cognitive model of analogical generalization, over the sets of known made decision for a specific culture. By doing so we can capture the norms and the commonalities in the decision making process of that culture. I discussed this process more in depth in Chapter 5, where I explored model construction from interview data for categorizing and making novel predictions about beliefs and norms of different cultures.

Second, I plan to compile story libraries for different cultural groups, based on my existing collaborations with cognitive psychologists and anthropologists. By gathering core cultural narratives of a certain culture and adding them to the KB of the system, we can investigate the effects of these narratives on moral decision making for that specific culture. Moreover, by adding story libraries for different cultures to the system we can model how these narratives can result in cross-cultural differences. My hope is that MoralDM can provide new insights about what is common, and what is different, about people's moral decision making across cultures.

Third, I plan to incorporate a cognitively plausible model of similarity-based retrieval, MAC/FAC (Forbus, Gentner, & Law, 1995), to make analogical reasoning more scalable as the story library grows. Once there are large number of stories in a library, running SME on every one of these stories is inefficient and not cognitively plausible. However, using MAC/FAC can make this process scalable by providing a robust method for similarity-based retrieval.

Fourth, I plan to add an emotion module to MoralDM. Bennis, Medin and Bartels (in press) argue that moral reasoning generally maps to the following two modes of decision making: recognition-based and affect-based. MoralDM in its current state implements recognition-based decision making. However, in order to model affect-based moral decision making we need to add an emotion module to our system. There are existing computational models of emotion such as EMA (Gratch & Marsella, 2004a) which have been incorporated in different systems for different purposes. For example, EMA has been incorporated in to a larger system for modeling emotions in virtual humans (Gratch & Marsella, 2004b). Adding an emotion module to MoralDM will have two benefits. First, we can then model affect-based decision making. Second, an emotion module will allow us to model comparisons across different types of moral goods. Currently, MoralDM can only handle tradeoffs within a single kind of moral good, e.g., lives against lives. However, many moral dilemmas involve trading one type of value for another type, such as a tradeoff between land and human life. I believe these types of comparisons require an emotion module as they appear to rely more heavily on affect-based reasoning. In these situations when one protected value is being traded for a different value, the emotional valance associated with each of these values and anticipated emotions resulting from their tradeoff need to be taken into account when making a decision. Cognitive Appraisal Theory (Ortony, Clore, & Collins, 1988; Frijda, 1986; Lazarus, 1991) which emphasizes tight couple between emotion, cognition and motivation may be able to account for these anticipated emotions. We plan to incorporate an implementation of this theory in MoralDM.

6.2. Cultural Narratives, Analogy and Moral Decision Making

Motivated by the results of MoralDM, in Chapter 4 I investigated the role of cultural narratives in moral decision making. Cultural narratives are often used to communicate moral values. As I argued before, these narratives and stories can be viewed as moral compasses for cultures, helping us distinguish moral actions from immoral ones. In the praise of "Books that Build Character" by Kilpatrick, Wolfe and Wolfe (1994), Amitai Etzioni writes: "Values do not fly on their wings. They are communicated, effectively, around stories, historical narratives, legends and such". I argued that many culturally specific and religious values are passed from generation to generation within cultural narratives. As a result, studying cultural narratives and more generally cultural products is essential to understanding the process of human decision making.

The second contribution of this dissertation was an examination of the use of analogical reasoning in understanding novel moral situations. In a series of experiments, I examined whether the processes by which core cultural narratives are applied in people's lives follow the principles of analogical retrieval and mapping. My experiments demonstrated how analogical accessibility and alignability influence the use of canonical moral narratives. I also showed that access to different moral stories can result in differences in moral preference across cultures. In conclusion, the results of my experiments suggest that analogical mapping from core cultural narratives can influence moral reasoning about current moral dilemmas. As predicted by the analogical account, the effects of the narratives discussed were seen only for Iranians, not for Americans, consistent with the claim that the effects stem from core narratives of the Iranian culture.

6.2.1. Future Work

In future, I plan to pursue several lines of follow-up studies. First, I plan to run the reverse study, using well known American moral narratives, such as the 'cannot tell a lie' narrative for George Washington, in experiments in Iran. Second, I plan to investigate how new sacred values emerge. In other words, how new sacred narratives and rhetoric (Marietta, 2008) can materialize protected values in people and change the mode of their decision making. Third, I am interested in investigating how sacred values diminish. For this purpose I plan to study decision making in addicts through different stages of addiction. Substance abusers are an excellent population to study for this purpose, because they start with the same values and obligations as other members of society, but as their addiction becomes more severe they tend to lose these values (and as they lose these values their addiction becomes even more severe).

6.3. Capturing Mental Models of Food Webs

In chapter 5, I introduced a modeling tool called QCM which enables cognitive scientists to build, simulate and explore qualitative mental models. This modeling tool offers a friendly interface for experimenters to explore causal models using QP theory semantics (Forbus, 1984). I demonstrated that QCM provides the formalism and the functionality necessary for modeling psychological data about causal mental models. Specifically, QCM was used to model transcripts of interview data conducted between two cultural groups. Next, a cognitively plausible classification technique, requiring order of magnitude smaller amount of training data than typical machine learning algorithms, was applied to classify the QCM models. I showed that cultural differences in causal reasoning about food webs can be captured to some degree in terms of similarities and differences in qualitative models extracted from transcript data. Qualitative representations proved to be an intuitive and valuable formalism for capturing these causal mental models. In conclusion, the results of my experiments confirmed the utility of using structured qualitative representations and analogical generalization for modeling causal mental models.

6.3.1. Future Work

I plan to extend QCM in several ways. First, I plan to incorporate the generalizations made from QCM models into MoralDM, by extracting first-principles rules from these generalizations. As I have shown in Section 4, causal mental models can be used to capture common properties and rules of decision making of a culture. These common properties can be thought of as structures and rules which dictate modes and types of decision making in that culture. Therefore, by extracting these rules from the generalization made from causal mental models and adding them to the first-principles reasoning module of MoralDM, we can capture some common behaviors in the decision making process of that culture.

I plan to use similarity-based qualitative simulation (Yan and Forbus 2005, 2004) to support creating predictions based on learned generalizations from transcript models. Standard qualitative simulation algorithms, although useful for many engineering applications, have some cognitively implausible properties. For example, the standard method for performing envisionment in qualitative reasoning works by exhaustively exploring every possibly world, and therefore does not match the flexibility, robustness, and speed of human reasoning (Yan and

Forbus 2005). Hybrid qualitative simulation (Forbus and Gentner 1997) which integrates similarity-based and first-principles reasoning would provide a more cognitively plausible account of human mental models reasoning.

I would like to use EA NLU (Tomai 2009a, Kuehne 2004) to semi-automatically construct formal representations in predicate calculus of the field-work transcripts. These predicate calculus statements can then be converted in to QCM models. This process may require extending the range of EA NLU coverage to handle transcripts of interview data. Automatic construction of QCM models from transcripts will expedite the process of model building. It will also make this process less error prone and it will reduce tailorability.

Open-ended interviews are useful for exploratory investigations of the ways in which participants are likely to respond to hypothetical scenarios, and future research can build on the knowledge gained here. Specifically, the present results can now be used as a basis for designing a participant-friendly interface for QCM in which participants themselves can engage in the process making QCM models. As discussed above, VModel, the predecessor of QCM, has been successfully used to teach middle-school students build qualitative models. Similarly, using tablets and touch-screen laptops, now widely available, an interface can be built in which participants can use a pen, or their fingers, to make models. This will, however, require changing the user-interface of QCM to provide a more intuitive method of model making. Just as CogSketch (Forbus, Usher, et al. 2008) is used to collect data about sketching (Jee, et al. 2009), QCM can then be used to collect data about the process of model building, such as timing data. Moreover, the interface could be used to present participants with a detailed list of probes. For, example a more comprehensive list of animals and plants that represent all of the trophic levels and ecological considerations mentioned, can easily be presented to participants. This feature should help provide a systematic probing of participants' knowledge and it should keep the overall size of the resultant models consistent across interviewees, without the extra steps of interviewing, transcribing and model building.

6.4. Final Thoughts

Computational modeling of cultural reasoning and decision making is receiving increasing attention in different communities. In cognitive sciences these models are generally used for better understanding of human cognition. In economics these models are applied to capture how different cultures perceive risks, probabilities and benefits. More recently, political scientists have shown interest in these modeling approaches for attempting to resolve cultural conflicts in a scientific manner. This dissertation provides the following major contributions to this modeling process. First, highly structural representations are essential for capturing distributed patterns of mental representations which constitute different aspects of culture. Second, integrating and applying cultural narratives in computational models of cultural reasoning are a prerequisite to understanding and explaining cultural behavior as these products are responsible for carrying the essence and the wisdom of a culture. As Roger Schank has famously argued, "knowing a culture means knowing the stories that the culture provides" (1995).

7. References

Allen, J. F. (1995). *Natural Language Understanding* (2nd ed ed.). Redwood City, CA: Benjamin/Cummings.

al-Naraqi, M. M. (18th century). *Jami 'al-Sa'adat. The Collector of Felicities.*

Anderson, M., Anderson, S., & Armen, C. (2006). An Approach to Computing Ethics. *IEEE Intelligent Systems , 21* (4), 56-63.

Åqvist, L. (1984). Deontic Logic. In D. Gabbay, & F. Guenthner (Eds.). Reidel, Dordrecht.

Ashby, D., & Smith, A. F. (2000). Evidence-based medicine as Bayesian decision-making. *Statistics in medicine , 19* (23), 3291-3305.

Atran, S. (1990). *Cognitive Foundations of Natural History.* Cambridge, England: Cambridge University Press.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences , 21 (4)*, 547-609.

Atran, S., Axelrod, R., & Davis, R. (2007). Sacred barriers to Conflict Resolution. *Science , 317*, 1039-1040.

Atran, S., Estin, P., Coley, J., & Medin, D. L. (1997). Generic species and basic levels: Essence and appearance in folkbiology. *Journal of Ethnobiology , 17*, 22-45.

Atran, S., Medin, D., & Ross, N. (2005). The cultural mind: Environmental decision making and cultural modeling within and across populations. *Psychological Review , 112*, 744-776.

Bang, M., Medin, D. L., & Atran, S. (2007). Cultural Mosaics and Mental Models of Nature. *Proceedings of the National Academy of Sciences , 104*, 13868-13874.

Bang, M., Townsend, J., Unsworth, S. J., & Medin, D. L. (2005). Cultural Models of Nature and Their Relevance to Science Education.

Barbour, I. G. (1974). *Myths, Models, and Paradigms: A Comparative Study in Science and Religion.* New York: Harper & Row.

Baron, J., & Spranca, M. (1997). Protected Values. *Organizational Behavior and Human Decision Processes , 70*, 1-16.

Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices? *Psychol Sci , 18* (1), 24-28.

Bennis, W. M., Medin, D. L., & Bartels, D. M. (in press). The Costs and Benefits of Calculation and Moral Rules. *Perspectives in Psychological Science* .

Berlin, B. (1992). *Ethnobiological Classification.* Princeton, NJ: Princeton University Press.

Berlin, B., Breedlove, D., & Raven, P. (1973). General Principles of Classification and Nomenclature in Folk Biology. *American Anthropologist , 75*, 214-242.

Berlin, B., Breedlove, D., & Raven, P. (1974). *Principles of Tzeltal plant classification; an introduction to the botanical ethnography of a Mayan-speaking people of highland Chiapas.* New York: Academic Press.

Biswas, G., Schwartz, D., Bransford, J., & Vanderbilt, T. T. (2001). Technology Support for Complex Problem Solving: From SAD Environments to AI. In K. Forbus, & P. Feltovich (Eds.). AAAI Press.

Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory and Cognition , 28*, 108-124.

Blockeel, H., & Uwents, W. (2004). Using Neural Networks for Relational Learning. *In Proceedings of ICML-2004 Workshop on Statistical Relational Learning and its Connection to Other Fields*, (pp. 23-28).

Bonet, B., & Geffner, H. (1996). Arguing for decisions: A qualitative model of decision making. *In Proc. of the 12th Conf. on Uncertainty in Artificial Intelligence (UAI'96* (pp. 98-105). Morgan Kaufmann.

Bouwer, A., & Bredeweg, B. (2001). VISIGARP: graphical representation of qualitative simulation models, in: Artificial Intelligence in Education. In J. D. Moore, G. L. Redfield, & J. L. Johnson (Eds.). Amsterdam, The Netherlands: IOS Press.

Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., & Marshall, M. S. (2001). GraphML Progress Report: Structural Layer Proposal. *Proceedings of the 9th International Symposium on Graph Drawing (GD-01).* Vienna, Austria.

Bredeweg, B., Bouwer, A., Jellema, J., Bertels, D., Linnebank, F., & Liem, J. (2007). Garp3 - A New Workbench for Qualitative Reasoning and Modelling. In D. Sleeman, & K. Barker (Ed.), *4th International Conference on Knowledge Capture (K-CAP 2007)*, (pp. 183-184). Whistler, BC, Canada.

Bredeweg, B., Salles, P., Bouwer, A., & Liem, J. (2006). Garp3 – A new Workbench for Qualitative Reasoning and Modelling. *Proceedings of the 20th International Workshop on Qualitative Reasoning (QR-06).*

Bringsjord, S., Arkoudas, K., & Bello, P. (2006). Toward a general logicist methodology for engineering ethically correct robots. *IEEE Intelligent Systems , 21 (4)*, 38-44.

Brown, C. (1984). *Language and Living Things: Uniformities in Folk Classification and Naming.* Brunswick, NJ: Rutgers University Press.

Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language Usage.* Cambridge, UK: Cambridge University Press.

Bruner, J. (1986). *Actual Minds, Possible Worlds.* Cambridge, MA: Harvard University Press.

Busemeyer, J. R., & Johnson, J. G. (2004). Computational models of decision making. In D. Koehler, & N. Harvey (Eds.). Oxford, UK: Blackwell.

Casasola, M. (2005). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development , 76 (1)*, 279-290.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science , 15*, 89-132.

Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline.* Cambridge, MA: Harvard University Press.

Colhoun, J., & Gentner, D. (2009). Drawing inferences from causal analogies. *In Proceedings of the 31th Annual Conference of the Cognitive Science Society (CogSci-09).*

Dague, P. (1994). Model-based Diagnosis of Analog Electronic Circuits. *Annals of Mathematics and Artificial Intelligence , 11*, 439-492.

Dague, P. (1993a). Numeric Reasoning with Relative Orders of Magnitude. *Proceedings of the 7th International Workshop on Qualitative Reasoning.*

Dague, P. (1993b). Symbolic Reasoning with Relative Orders of Magnitude. *Proceedings of the 13th International Joint Conference of Artificial Intelligence (IJCAI-93).*

Dague, P., Deves, P., & Raiman, O. (1987). Troubleshooting: when Modeling is the Trouble. *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence.*

Darwiche, A. (2001). Recursive Conditioning. *Artificial Intelligence , 126* (1-2), 5-41.

Dayanik, A., & Nevill-Manning, C. G. (2004). Clustering in Relational Biological Data. *In Proceedings of ICML-2004 Workshop on Statistical Relational Learning and Connections to Other Fields*, (pp. 42-47).

Dehghani, M., & Forbus, K. (2009). QCM: A QP-Based Concept Map System. *In Proceedings of the 23nd International Workshop on Qualitative Reasoning*. Ljubljana, Slovenia.

Dehghani, M., Gentner, D., Forbus, K., Ekhtiari, H., & Sachdeva, S. (2009). Analogy and Moral Decision Making. *Analogy09*. Sofia, Bulgaria

Dehghani, M., & Lovett, A. (2006). Efficient Genre Classification using Qualitative Representations. *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR-06).* Victoria, Canada.

Dehghani, M., Sachdeva, S., Ekhtiari, H., Gentner, D., & Forbus, K. (2009). The Role of Cultural Narratives in Moral Decision Making. *In Proceedings of the 31th Annual Conference of the Cognitive Science Society*. Amsterdam, Holland.

Dehghani, M., Tomai, E., Forbus, K., & Klenk, M. (2008). An Integrated Reasoning Approach to Moral Decision-Making. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence* (AAAI-08). Chicago, IL.

Dehghani, M., Tomai, E., Forbus, K., & Klenk, M. (2008b). Order of Magnitude Reasoning in Modeling Moral Decision-Making. *Proceedings of the 22nd International Workshop on Qualitative Reasoning (QR-08)*. Boulder, CO.

Dehghani, M., Tomai, E., Forbus, K., Iliev, R., & Klenk, M. (2008a). MoralDM: A Computational Modal of Moral Decision-Making. *Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci-08)*. Washington, DC.

Dehghani, M., Tomai, E., Forbus, K., Iliev, R., & Klenk, M. (2008a). MoralDM: A Computational Modal of Moral Decision-Making. 2008 meeting of Society of Judge and Decision Making (SJDM-08). Chicago, IL.

Dehghani, M., Unsworth, S., Lovett, A., & Forbus, K. (2007). Capturing and Categorizing Mental Models of Food Webs using QCM. Proceedings of the 21st International Workshop on Qualitative Reasoning (QR-07). Aberystwyth, U.K.

Dickerson, J., Martinez, M. V., Reforgiato, D., & Subrahmanian, V. S. (2008). CIG: Cultural Island and Games. In V. S. Subrahamanian, & A. Kruglanski (Ed.), *Proceedings of the Second International Conference on Computational Cultural Dynamics (ICCCD-2008).* College Park, Maryland: AAAI Press.

Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review , 106*, 180-209.

Dubois, D., Godo, L., Prade, H., & Zapico, A. (1998). Making decision in a qualitative setting: From decision under uncertainty to case-based decision. *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, (pp. 594-605).

Edwards, W. (1954). The Theory of Decision Making. *Psychological Bulletin , 41*, 380-417.

Ekhtiari, H., Behzadi, A., Dehghani, M., Jannati, A., & Mokri, A. (in press). Prefer a cash slap in your face over credit for halva. *Journal of Judgment and Decision Making* .

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithms and Examples. *Artificial Intelligence , 41*, 1-63.

Fillmore, C. (1982). Frame semantics. Seoul: Hanshin.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology , 18*, 253-292.

Forbus, K. (1984). Qualitative Process Theory. *Artificial Intelligence , 24*, 85-168.

Forbus, K. (1996). Qualitative Reasoning. CRC Press.

Forbus, K., & Gentner, D. (1997). Qualitative mental models: Simulations or memories? *Proceedings of the Eleventh International Workshop on Qualitative Reasoning.*

Forbus, K., & Oblinger, D. (1990). Making SME Greedy and Pragmatic. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society (CogSci-1990).*

Forbus, K., Carney, K., Harris, R., & Sherin, B. (2001). A qualitative modeling environment for middle-school students: A progress report. *Proceedings of the 15th International Workshop on Qualitative Reasoning.*

Forbus, K., Carney, K., Sherin, B., & Ureel, L. (2004). Qualitative modeling for middle-school students. *Proceedings of the 18th International Qualitative Reasoning Workshop (QR-04).* Evanston, Illinois.

Forbus, K., Ferguson, R., & Gentner, D. (1994). Incremental Structure-Mapping. *Proceedings of the 16th Annual Conference of the Cognitive Science Society (CogSci-1994).*

Forbus, K., Gentner, D., & Law, K. (1995). MAC/FAC: A model of Similarity-based Retrieval. *Cognitive Science , 19* (2), 141-205.

Forbus, K., Gentner, D., & Law, K. (1995). MAC/FAC: A model of Similarity-based Retrieval. *Cognitive Science , 19 (2)*, 141-205.

Forbus, K., Usher, J., & Tomai, E. (2005). Analogical Learning of Visual/Conceptual Relationships in Sketches. *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05).*

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. *In the Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling.* Annecy, France.

Friedman, S. E., & Forbus, K. D. (2008). Learning Causal Models via Progressive Alignment & Qualitative Modeling: A Simulation. *In the Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci), Washington, D.C.*

Friedman, S., Taylor, J., & Forbus, K. (2009). Learning Naive Physics Models by Analogical Generalization. *In Proceedings of the 2nd International Analogy Conference.* Sofia, Bulgaria.

Frijda, N. (1986). *The Emotions. Studies in Emotion and Social Interaction.* New York, NY: Cambridge University Press.

Gelman, R., & Wellman, H. M. (1991). Insides and essences. *Cognition , 38*, 214-244.

Gelman, S. A. (2003). *The essential child: origins of essentialism in everyday thought.* Oxford University Press.

Gentner, D. (1988). Metaphor as structure mapping: the relational shift. *Child Development , 59*, 47-59.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science , 7* (2), 155-170.

Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. In E. Amsel, & J. P. Byrnes (Eds.). Lawrence Erlbaum Associates.

Gentner, D., & Markman, A. B. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist , 52*, 45-56.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development , 14*, 487-513.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. Gelman, & J. Byrnes (Eds.). Cambridge University Press.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology , 95(2)*, 393-408.

Gentner, D., Rattermann, M. J., & Forbus, K. (1993). The Roles of Similarity in Transfer: Separating Retrievability from Inferential Soundness. *Cognitive Psychology , 25*, 524-575.

Getoor, L., Koller, N. F., & Pfeffer, A. (2001). Learning Probabilistic Relational Models. In S. Dzeoski, & L. N (Eds.). Springer-Verlag.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology , 15*, 1-38.

Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development.* Cambridge, MA: Harvard University Press.

Glad, B., & Taber, C. S. (1990). Images, learning and the decision to use force: The domino theory of the United States. In B. Glad (Ed.). Sage Publications.

Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.). San Diego: Academic Press.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology , 23*, 222-262.

Gratch, J., & Marsella, S. (2004a). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research , 5*, 269-306.

Gratch, J., & Marsella, S. (2004b). Evaluating the modeling and use of emotion in virtual humans. In N. R. Jennings, C. Sierra, L. Sonenberg, & M. Tambe (Ed.), *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS 2004)* (pp. 320-?327). New York: ACM Press.

Gregan-Paxton, J. (2001). The role of abstract and specific knowledge in the formation of product judgments: an analogical learning perspective. *Journal of Consumer Psychology , 11(3)*, 141-158.

Guarini, M. (2005). Particularism and generalism: how AI can help us to better understand moral cognition.

Guo, F. Y., & Holyoak, K. J. (2002). Understanding similarity in choice behavior: A connectionist model. *Proceedings of the Cognitive Science Society Meeting.*

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture and morality, or, is it wrong to eat your dog? *Journal of Personality and Social Psychology , 65*, 613-628.

Halstead, D., & Forbus, K. (2007). Some Effects of a Reduced Relational Vocabulary on the Whodunit Problem. *In Proceedings of the 20th International Joint Conference on Artificial Intelligence.* Hyderabad, India.

Halstead, D., & Forbus, K. (2005). Transforming between propositions and features: Bridging the gap. *Proceedings of the 25th National Conference on Artificial Intelligence.*

Hasan, A. (2007). *Analogical Reasoning in Islamic Jurisprudence.* New Delhi, India: Adam Publishers and Distributors.

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology , 52*, 653-683.

Hays, T. (1983). Ndumba folkbiology and general principles of ethnobotanical classification and nomenclature. *American Anthropologist , 85*, 592-611.

Hilpinen, R. (2001). Deontic Logic. In L. Goble (Ed.). Blackwell.

Hohfeld, W. N. (1913). Fundamental legal conceptions as applied to judicial reasoning. *Yale Law Journal , 23*, 16-59.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition , 15*, 332-340.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology , 128*, 3-31.

Horty, J. (2001). *Agency and Deontic Logic.* Oxford University Press.

Hsee, C., & Weber, E. (1997). A Fundamental Prediction Error: Self-Others Discrepancies in Risk Preference. *Journal of Experimental Psychology: General , 126 (1)*, 45-53.

Hsee, C., & Weber, E. (1999). National Differences in Risk Preference and Lay Predictions. *Journal of Behavioral Decision Making , 12*, 165-179.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research , 9*, 90-98.

Hunn, E. (1977). *Tzeltal Folk Zoology: The Classification of Discontinuities in Nature.* New York: Academic Press.

Iliev, R., Sachdeva, S., Bartels, D. M., Joseph, C., Suzuki, S., & Medin, D. L. (2009). Attending to Moral Values. In D. Bartels, D. L. Medin, C. W. Bauman, B. H. Ross, & L. J. Skitka (Eds.). Elsevier. Psychology of Learning and Motivation, Vol. 50: Moral Judgment and Decision Making. San Diego, CA: Academic Press.

Jee, B., Gentner, D., Forbus, K., Sageman, B., & Uttal, D. (2009). Drawing on Experience: Use of Sketching to Evaluate Knowledge of Spatial Scientific Concepts. *In the Proceedings of the 31st Annual Conference of the Cognitive Science Society (CogSci-09).* Amsterdam, Holland.

Johnson, D. F., & Goldman, S. R. (1987). Children's Recognition and use of Rules of Moral Conduct in Stories. *American Journal of Psychology , 100*, 205-224.

Joyce, J. M. (1999). *The foundations of causal decision theory.* Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of decision under risk. *American Psychologist , 39* (4), 341-350.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge and New York: Cambridge University Press.

Kamp, H., & Reyle, U. (1993). *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory* (Vol. 42). Springer.

Kant, I. (1781/2003). *Critique of Pure Reason.* New York: Dover.

Keppens, J. (2007a). On Qualitative Probabilities for Legal Reasoning about Evidence. *Proceedings of the 21st International Workshop on Qualitative Reasoning.*

Keppens, J. (2007b). Towards Qualitative Approaches to Bayesian Evidential Reasoning. *Proceedings of the 11th International Conference on Artificial Intelligence and Law.*

Kerr, N. L. (1993). Stochastic models of juror decision making. In R. Hastie (Ed.). New York: Cambridge University Press.

Kilpatrick, W. (1992). *Why Johnny Can't Tell Right from Wrong and What We Can Do About It.* Touchstone.

Kilpatrick, W., Wolfe, G., & Wolfe, S. M. (1994). *Books That Build Character: A Guide to Teaching Your Child Moral Values Through Stories.* Touchstone.

Kohlberg, L. (1981). *Essays on Moral Development: Vol. 1. The Philosophy of Moral Development.* New York: Harper & Row.

Kohlberg, L. (1984). *Essays on Moral Development: Vol. 2. The Psychology of Moral Development.* New York: Harper & Row.

Kokinov, B. (2005). Can a Single Episode or a Single Story Change our Willingness to Risk? The Role of Analogies in Decision-Making. *Advances in Cognitive Economics* .

Kruglanski, A. W. (1989). The psychology of being "right": the problem of accuracy in social perception of cognition. *psychological bulletin , 106*, 395-409.

Kuehne, S. (2004). *Understanding natural language descriptions of physical phenomena.* Ph.D. Thesis, Department of Computer Science. Northwestern University.

Kuehne, S., & Forbus, K. (2004). Capturing QP-Relevant Information from Natural Language Text. *In Proceedings of the 18th International Workshop on Qualitative Reasoning (QR04).*

Kuehne, S., Forbus, K., Gentner, D., & Quinn, B. (2000). SEQL: Category Learning as Progressive Abstraction Using Structure Mapping. *In Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, (pp. 770-775). Philadelphia, PA.

Kuhmerker, L. (1975). Learning to Care: The Development of Empathy. *Journal of Moral Education , 5*, 25.

Lacoste, J.-Y. (2004). *Encyclopedia of Christian Theology.* Routledge.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *psychological science , 1*, 362-370.

Lazarus, R. S. (1991). *Emotion and Adaptation.* New York, NY: Oxford University Press.

Lim, C. S., & Baron, J. (1997). *Protected values in Malaysia Singapore, and the United States.* Department of Psychology, University of Pennyslvania.

Lockwood, J. H. (1996). *The Moral of the Story: Content, Process, and Reflection in Moral Education Through Narratives.* Ph.D. Thesis, University of Florida.

Lockwood, K., Lovett, A., & Forbus, K. (2008). Automatic Classification of Containment and Support Spatial Relations in English and Dutch. *Proceedings of the international conference on Spatial Cognition VI* (pp. 283-294). Berlin, Heidelberg: Springer-Verlag.

Lovett, A., Dehghani, M., & Forbus, K. (2007). Incremental Learning of Perceptual Categories for Open-Domain Sketch Recognition. *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07).* Hyderabad, India: AAAI Press.

Lyotard, J.-F. (1984). *The Postmodern Condition: A Report on Knowledge.* Minneapolis, MN: University of Minnesota Press.

Machado, V., & Bredeweg, B. (2001). Towards Interactive Tools for Constructing Articulate Simulations. *Proceedings of the 15th International workshop on Qualitative Reasoning, (QR-01)*, (pp. 98-104). San Antonio, Texas.

MacIntyre, A. (1981). *After Virtue: A Study in Moral Theory.* Notre Dame, IN: University of Notre Dame Press.

Macleod, C., Grishman, R., & Meyers, A. (1998). COMLEX Syntax Reference Manual, Version 3.0.

Marietta, M. (2008). From My Cold, Dead Hands: Democratic Consequences of Sacred Rhetoric. *The Journal of Politics , 70 (3)*, 767-?779.

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science , 21(4)*, 373-418.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology , 25*, 431-467.

Markman, A. B., & Medin, D. L. (2002). Decision Making. In D. L. Medin, & H. Pashler (Eds.). New York: John Wiley and Sons.

Markman, A. B., & Moreau, C. P. (2001). Analogy and analogical comparison in choice. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.). The MIT Press.

Martinez, M. V., Simari, G. I., Sliva, A., & Subrahmanian, V. S. (2008). CONVEX: Context Vectors as a Similarity-based Paradigm for Forecasting Group Behaviors. *IEEE Intelligent Systems , 23*, 51-57.

Mavrovouniotis, M. L., & Stephanopoulos, G. (1990). Formal order-of-magnitude reasoning in process engineering. In D. S. Weld, & J. d. Kleer (Eds.). Morgan Kaufmann Publishers Inc.

Mavrovouniotis, M. L., & Stephanopoulos, G. (1988). Order of Magnitude Reasoning in Process Engineering. *Computers & Chemical Engineering , 12*.

Mavrovouniotis, M. L., & Stephanopoulos, G. (1987). Reasoning with orders of magnitude and approximate relations. *Proceedings of the Sixth National Conference on Artificial Intelligence.*

May, E. R. (1973). *"Lessons" of the Past.* New York, NY: Oxford University Press.

McLaren, B. M. (2006). Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions. *IEEE Intelligent Systems , July/August*, 29-37.

Medin, D. L., & Atran, S. (2004). The Native Mind: Biological Categorization and Reasoning in Development and Across Cultures. *Psychological Review , 11(4)*, 960-983.

Medin, D. L., Goldstone, R. L., & Markman, A. B. (1995). Comparison and choice: Relations between similarity processing and decision processing. *Psychonomic Bulletin and Review , 2*, 1-19.

Medin, D. L., Ross, N. O., Atran, S., Burnett, R., & Blok, S. (2002). *Categorization and Reasoning in Relation to Culture and Expertise.* New York: Academic Press.

Mellers, B. A., Schwartz, A., & Cooke, A. D. (1998). Judgment and Decision Making. *Annual Review of Psychology , 49*, 447-477.

Meyer, J. -J., Dignum, F. P., & Wieringa, R. J. (1994). *The Paradoxes of Deontic Logic Revisited: A Computer Science Perspective. Or: Should computer scientists be bothered by the concerns of philosophers?* The Netherlands: Utrecht University.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences , 11 (4)*, 143-152.

Miller, C., Wu, P., Funk, H., & Johnson, W. L. (2007). A Computation Approach to Etiquette and Politeness: An Etiquette Engine for Cultural Interaction Training. *In Proceedings of Behavior Representation in Modeling & Simulation (BRIMS-07).* Norfolk, VA.

Murakami, Y. (2004). Utilitarian Deontic Logic. *Proceedings of the 5th International Conference Advances in Modal Logic*, (pp. 288-302).

Musavi-Lari, M. (1997). *Ethics and spiritual growth.* Qum, Iran: Foundation of Islamic Cultural Propagation in the World.

Narvaez, D. (2002). Does Reading Moral Stories Build Character? *Educational Psychology Review , 14 (2)*, 155-171.

Narvaez, D. (2005). The Neo-Kohlbergian Tradition and Beyond: Schemas, Expertise and Character. In C. Pope-Edwards, & G. Carlo (Eds.). Lincoln, NE: University of Nebraska Press.

Narvaez, D., Bentley, J., Gleason, T., & Samuels, J. (1998). Moral theme comprehension in third grade, fifth grade and college students. *Reading Psychology , 19(2)*, 217-?241.

Narvaez, D., Gleason, T., Mitchell, C., & Bentley, J. (1999). Moral Theme Comprehension in Children. *Journal of Educational Psychology , 91 (3)*, 477-487.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule based classification learning: A replication and extension of Shepard, Hovland, and Jenkins. *Memory and Cognition , 22*, 1961.

Novak, J. D., & Gowin, D. B. (1984). *Learning How To Learn.* New York: Cambridge University Press.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, memory, and Cognition , 14*, 510-520.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions.* Cambridge: Cambridge University Press.

Parducci, A., & Perret, L. (1971). Category-Rating Scales: Effects of Relative Spacing and Frequency of Stimulus Values. *Journal of Experimental Psychology , 89* (2), 427-452.

Parmigiani, G. (2002). *Modeling in Medical Decision Making: A Bayesian Approach.* UK: Wiley.

Parsons, S. (1995). Refining reasoning in qualitative probabilistic networks. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence.*

Parsons, S., & Mamdani, E. (1993). On reasoning in networks with qualitative uncertainty. *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence.*

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral Decision Research: A Constructive Processing Perspective. *Annual Review of Psychology , 43*, 87-131.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference.* London: Cambridge University Press.

Pearl, J. (1993). From conditional oughts to qualitative decision theory. *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93).*

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* San Francisco: Morgan Kaufmann.

Pennington, N., & Hastie, R. (1993). The story model of juror decision making. In R. Hastie (Ed.). New York: Cambridge University Press.

Pereira, L. M., & Saptawijaya, A. (2007). Modelling Morality with Prospective Logic. *EPIA Workshops 2007*, (pp. 99-11).

Petkov, G., & Kokinov, B. (2006). JUDGEMAP - Integration of Analogy-Making, Judgment, and Choice. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Petrov, A., & Anderson, J. (2005). The Dynamics of Scaling: A Memory-Based Anchor Model of Category Rating and Absolute Identification. *Psychological Review , 112* (2), 383-416.

Piaget, J. (1932/1965). *The Moral Judgment of the Child.* New York: Free Press.

Plumeley, M. D. (1994). *Towards Optimal Learning from Very Small Data Sets.* London, UK: Department of Computer Science, King's College.

Powers, T. M. (2006). Prospects for a Kantian machine. *IEEE Intelligent Systems , 21 (4)*, 46-51.

Prasad, L. (2007). *Poetics of Conduct.* New York: Columbia University Press.

Raiman, O. (1991). Order of Magnitude Reasoning. *Artificial Intelligence , 51*, 11-38.

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin , 115 (3)*, 381-400.

Ritov, I., & Baron, J. (1999). Protected Values and Omission Bias. *Organizational Behavior and Human Decision Processes , 79* (2), 79-94.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology , 7(4)*, 573-605.

Ross, B. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition , 15*, 456-468.

Ross, N. O., Medin, D. L., Coley, J. D., & Atran, S. (2003). Cultural and experimental differences in the development of folkbiological induction. *Cognitive Development , 18(1)*, 25-47.

Rzepka, R., & Araki, K. (2005). What could statistics do for ethics? The Idea of Common Sense Processing Based Safety Valve.

Schank, R. (1995). *Tell Me a Story: Narrative and Intelligence.* Northwestern University Press.

Shutters, S. T., & Cutts, B. B. (2008). A Simulation Model of Cultural Consensus and Persistent Conflict. In V. S. Subrahamanian, & A. Kruglanski (Ed.), *Proceedings of the Second International Conference on Computational Cultural Dynamics (ICCCD-2008)* (pp. 71-78). College Park, Maryland: AAAI Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In A. M. Brandt (Ed.). New York: Routledge Press.

Slade, S. (1994). *Goal-based decision making: An interpersonal model.* Hillsdale, NJ: Lawrence Erlbaum.

Sousa, P., Atran, S., & Medin, D. (2002). Essentialism and folkbiology: Evidence from Brazil. *Journal of Cognition and Culture , 2 (3)*, 195-223.

Sox, H. C., Blatt, M. A., Higgins, M. C., & Marton, K. I. (2007). *Medical Decision Making.* ACP Press.

Sperber, D. (1985). Anthropology and Psychology: Towards an epidemiology of representations. *Man , 20*, 73-89.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach.* Oxford: Blackwell.

Subrahmanian, V. S., Albanese, M., Martinez, M. V., Nau, D., Reforgiato, D., Simari, G. I., et al. (2007). CARA: A Cultural Adversarial Reasoning Architecture. *IEEE Intelligent Systems , 22* (2), 12-16.

Tengli, A., Dubrawski, A., & Chen, L. (2005). Predictive Rule Lists for Modeling Small Sets of Dirty Data. *In the Proceedings of IEEE International Conference on Information and Automation.* Colombo, Sri Lanka.

Tetlock, P. E. (2000). Cognitive biases and organizational correctives: Do both disease and cure depend on the ideological beholder? *Administrative Science Quarterly , 45* (2), 293-326.

Tetlock, P. E. (2003). Thinking the Unthinkable: Sacred Values and Taboo Cognitions. *Trends in Cognitive Sciences , 7*, 320-324.

Thagard, P., & Millgram, E. (1995). Inference to the best plan: A coherence theory of decision. In A. Ram, & D. B. Leake (Eds.). MIT Press.

Tomai, E. (2009a). *A pragmatic approach to computational narrative understanding.* Ph.D. Thesis, Department of Electrical Engineering and Computer Science. Northwestern University.

Tomai, E., & Forbus, K. D. (2009b). EA NLU: Practical Language Understanding for Cognitive Modeling. *In Proceedings of the Twenty-Second International FLAIRS Conference.*

Tomai, E., & Forbus, K. D. (2008). Using Qualitative Reasoning for the Attribution of Moral Responsibility. *In the Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci), Washington, D.C.*

Tversky, A. (1972). Elimination by aspects: A theory of choice. *psychological review* , 281-299.

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics , 106* (4), 1039-1061.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: A model based on principles of neural computation. *Psychological Review , 108*, 550-592.

Uttal, D. H., Gentner, D., Liu, L. L., & Lewis, A. R. (2008). Developmental changes in children's understanding of the similarity between photographs and their referents. *Developmental Science , 11(1)*, 156-170.

Vitz, P. C. (1990). The Use of Stories in Moral Development: New Psychological Reasons for an Old Education Method. *American Psychologist , 45 (6)*, 709-720.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Waldmann, M. R., & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science , 18* (3), 247-253.

Weber, E. U. (1998). *From Shakespeare to Spielberg: Predicting modes of decision making. Presidential Address, Annual Meeting, Society of Judgment and Decision Making, Dallas, TX.*

Weber, E. U., & Hsee, C. K. (1998). Cross-cultural Differences in Risk Perception, but Cross-cultural Similarities in Attitudes Towards Perceived Risk. *Management Science , 44 (9)*, 1205-1214.

Weber, E. U., & Hsee, C. K. (2000). Culture and Individual Judgment and Decision Making. *Journal of Applied Psychology , 49*, 32-61.

Weber, E. U., Ames, D., & Blais, A. R. (2005). "How do I choose thee? Let me count the ways:" A textual analysis of similarities and differences in modes of decision making in China and the United States. *Management and Organization Review , 1*, 87-118.

Wellman, M. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence , 44*, 257-303.

Wilder, P. G. (1980). The Moral of a Story: Preschoolers' Gradual Comprehension of a Narrative on Sesame Street. *Moral Education Forum , 5*, 2-14.

Wilson, N. (1995). An order of magnitude calculus. *Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-95).*

Winston, J. (1998). *Drama, Narrative and Moral Education.* Bristol, PA: Falmer Press, Taylor & Francis Inc.

Winston, P. H. (1982). Learning New Principles from Precedents and Exercises. *Artificial Intelligence , 19(3)*, 321-350.

Wright, G. v. (1951). Deontic logic. *Mind , 60*, 1-15.

Yan, J., & Forbus, K. (2005). Similarity-based qualitative simulation. *In Proceedings of the 27th Annual Meeting of the Cognitive Science Society (CogSci-05).*

Yan, J., & Forbus, K. (2004). Similarity-based qualitative simulation: A preliminary report. *In Proceedings of the 18th International Qualitative Reasoning Workshop (QR-04).* Evanston, Illinois, USA.

Yates, J. F., Lee, J. W., & Bush, J. G. (1997). General knowledge overconfidence: Cross-national variations, response style, and "reality". *Organizational Behavior and Human Decision Processes , 70 (2)*, 87-94.

Yates, J. F., Lee, J. W., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes , 65 (2)*, 138-147.

Yates, J. F., Zhu, Y., Ronis, D. L., Wang, D. F., Shinotsuka, H., & Toda, W. (1989). Probability judgment accuracy: China, Japan, and the United States. *Organizational Behavior and Human Decision Processes , 43 (2)*, 145-171.

8. Appendix

8.1. Appendix A: Sample Scenarios and their Representations

8.1.1. Starvation Scenario

The original scenario from Ritov and Baron (1999):

> A convoy of food trucks is on its way to a refugee camp during a famine in Africa. (Airplanes cannot be used.) You find that a second camp has even more refugees. If you tell the convoy to go to the second camp instead of the first, you will save 1000 people from death, but 100 people in the first camp will die as a result.

The simplified English version of the scenario used as input to EA NLU:

> A convoy of trucks is transporting food to a refugee camp during a famine in Africa. 1000 people in a second refugee camp will die. You can save them by ordering the convoy to go to that refugee camp. The order will cause 100 people to die in the first refugee camp.

Predicate calculus representation of the above scenario produced by EA NLU:

```
'(FAMINE-1
  (isa Inaction131950 Inaction)
  (isa Inaction131950 (InactionFn order131049))
  (performedBy PreventingSomething131948 you128898)
  (preventedSit PreventingSomething131948 die128829)
  (isa PreventingSomething131948 PreventingSomething)
  (isa SelectingSomething131949 SelectingSomething)
  (choices SelectingSomething131949 order131049)
  (choices SelectingSomething131949 Inaction131950)
  (causes-PropSit (chosenItem SelectingSomething131949 Inaction131950) die128829)
  (causes-PropSit (chosenItem SelectingSomething131949 order131049) save128937)
  (isa convoy127246 Convoy)
  (isa die128829 Dying)
  (relationInstanceMember objectOfStateChange die128829 them128970)
  (in-UnderspecifiedContainer die131270 refugee-camp129739)
  (isa die131270 Dying)
  (relationInstanceMember objectOfStateChange die131270 set-of-people131188)
  (isa famine127520 Famine)
  (in-UnderspecifiedContainer famine127520 ContinentOfAfrica)
```

```
(isa food127320 FoodOrDrink)
(performedBy order131049 you128898)
(recipientOfInfo order131049 convoy127246)
(infoTransferred order131049
   (and (isa refugee-camp129739 RefugeeCamp)
        (isa convoy127246 Convoy)
        (primaryObjectMoving go129115 convoy127246)
         (toLocation go129115 refugee-camp129739)
         (isa go129115 Movement-TranslationEvent)))
(causes-SitProp order131049
   (and (isa die131270 Dying)
         (in-UnderspecifiedContainer die131270 refugee-camp129739)
         (isa set-of-people131188 Set-Mathematical)
         (cardinality set-of-people131188 100)
         (relationInstanceMember objectOfStateChange die131270 set-of-people131188)
         (relationMemberInstance isa set-of-people131188 Person)
        (isa refugee-camp129739 RefugeeCamp)))
(isa order131049 Ordering-CommunicationAct)
(isa refugee-camp127421 RefugeeCamp)
(isa refugee-camp129739 RefugeeCamp)
(isa save128937 RescuingSomeone)
(beneficiary save128937 them128970)
(performedBy save128937 you128898)
(by-Underspecified save128937 order131049)
(causes-SitSit save128937 PreventingSomething131948)
(cardinality set-of-people131188 100)
(isa set-of-people131188 Set-Mathematical)
(relationMemberInstance isa set-of-people131188 Person)
(relationMemberInstance isa them128970 Person)
(relationMemberInstance in-UnderspecifiedContainer them128970 refugee-camp129739)
(cardinality them128970 1000)
(isa them128970 Set-Mathematical)
(to-UnderspecifiedLocation transport127299 refugee-camp127421)
(transporter transport127299 convoy127246)
(transportees transport127299 food127320)
(isa transport127299 TransportationEvent)
(isa truck127274 Truck)
(possessiveRelation truck127274 convoy127246))
```

8.1.2. Financial Assistance Scenario

The original scenario from Ritov and Baron (1999):

> Your office provides financial assistance to a plant employing 50 workers. If you withdraw this support (which will put 50 workers out of work) you can use the funds to support another plant, which employs 500 workers. Without government support, this second plant will close down.

The simplified English version of the scenario used as input to EA NLU:

> Your office provides financial support to a plant that employs 50 workers. Another plant employs 500 workers. That plant will close. The closing will cause the 500 workers to be unemployed. You can save them by transferring the financial support. The transfer will cause the 50 workers to be unemployed.

Predicate calculus representation of the above scenario produced by EA NLU:

```
'(FINAN-1
   (isa Inaction91624 Inaction)
   (isa Inaction91624 (InactionFn transfer91105))
   (performedBy PreventingSomething91622 you90799)
   (preventedProp PreventingSomething91622
       (and (isa set-of-worker90320 Set-Mathematical)
       (cardinality set-of-worker90320 500)
       (relationMemberInstance hasAttributes set-of-worker90320 UnemployedPerson)
       (relationMemberInstance isa set-of-worker90320 Employee)))
   (isa PreventingSomething91622 PreventingSomething)
   (isa SelectingSomething91623 SelectingSomething)
   (choices SelectingSomething91623 transfer91105)
   (choices SelectingSomething91623 Inaction91624)
   (causes-PropProp (chosenItem SelectingSomething91623 Inaction91624)
       (and (isa set-of-worker90320 Set-Mathematical)
            (cardinality set-of-worker90320 500)
            (relationMemberInstance hasAttributes set-of-worker90320 UnemployedPerson)
            (relationMemberInstance isa set-of-worker90320 Employee)))
   (causes-PropSit (chosenItem SelectingSomething91623 transfer91105) save90838)

   (causes-SitProp close90499
       (and (isa set-of-worker90320 Set-Mathematical)
            (cardinality set-of-worker90320 500)
            (relationMemberInstance hasAttributes set-of-worker90320 UnemployedPerson)
            (relationMemberInstance isa set-of-worker90320 Employee)))
   (isa close90499 ClosingSomething)
   (isa financial-support89816 FinancialSupport)
   (isa office89775 OfficeSpace)
   (isa plant89874 FactoryBuildingComplex)
   (relationInstanceMember hasWorkers plant89874 set-of-worker90031)
   (relationInstanceMember hasWorkers plant90251 set-of-worker90320)
   (isa plant90251 FactoryBuildingComplex)
   (target provide89786 plant89874)
   (transferredObject provide89786 financial-support89816)
   (performedBy provide89786 office89775)
   (isa provide89786 MakingSomethingAvailable)
   (to-UnderspecifiedLocation provide89786 plant89874)
   (isa save90838 RescuingSomeone)
   (beneficiary save90838 set-of-worker90320)
   (performedBy save90838 you90799)
   (by-Underspecified save90838 transfer91105)
   (causes-SitSit save90838 PreventingSomething91622)
   (cardinality set-of-worker90031 50)
   (isa set-of-worker90031 Set-Mathematical)
   (relationMemberInstance hasAttributes set-of-worker90031 UnemployedPerson)
   (relationMemberInstance isa set-of-worker90031 Employee)
   (cardinality set-of-worker90320 500)
   (isa set-of-worker90320 Set-Mathematical)
   (relationMemberInstance hasAttributes set-of-worker90320 UnemployedPerson)
   (relationMemberInstance isa set-of-worker90320 Employee)
   (performedBy transfer91105 you90799)
   (transferredObject transfer91105 financial-support89816)
   (causes-SitProp transfer91105
      (and (isa set-of-worker90031 Set-Mathematical)
           (cardinality set-of-worker90031 50)
           (relationMemberInstance hasAttributes set-of-worker90031 UnemployedPerson)
           (relationMemberInstance isa set-of-worker90031 Employee)))
   (isa transfer91105 GeneralizedTransfer))
```

As discussed in section 3.2, after a case is solved, the decision chosen along with reasons for choosing that decision and the mode in which the decision was made in are stored with the case itself in the case library for future use. The additional statements added to the case are either derived from the rules of the FPR module, or are information mapped from the base analog. The following statements are the additional information saved with the above scenario:

```
(implies
    (and (involvesSacredValue Inaction91624)
         (involvesSacredValue transfer91105)
         (notOrdersOfMagnitudeDifferent SelectingSomething91623)
         (isInaction Inaction91624))
    (pureDeontologicalChoice Inaction91624))
(makeDecision Inaction91624)
```

8.1.3. Pouria Vali Scenario

The original scenario:

The original scenario from Ritov and Baron (1999):

> Pourya Vali was the most famous wrestler of his time. The morning before wrestling with a young athlete from another province, he goes to a mosque and sees the mother of the young athlete praying and saying "God, my son is going to wrestle with Pourya Vali. Please watch over him and help him win the match so he can use the prize money to buy a house". Pourya Vali thinks to himself that the young wrestler needs the money more than he does, and also winning the match will break the heart of the old mother. He has two choices, he can either win the match and keep his status as the best wrestler in the world or he could lose the match and make the old mother happy. Even though he was known not to ever lose a match, he loses that one on purpose.

The simplified English version of the scenario used as input to EA NLU:

> Pouryaie Vali was the most famous wrestler of his time. He was going to wrestle a young athlete from another province. He goes to a mosque and sees the mother of the young athlete praying. She says, "My son is going to wrestle Pouryaie Vali. Please help him to win the match so that he can use the prize money to buy a house." Pouryaie Vali thinks that the young athlete needs the money more than he does. He also thinks that winning the match will break the old mother's heart. He has two options. The first option is, he can win the match. This would keep his status as the

best wrestler. The second option is, he can lose the match to make the old mother happy. This would risk his status and help the young athlete to buy a house. He makes the old mother happy by choosing the second option.

Predicate calculus representation of the above scenario produced by EA NLU:

```
'(vali-base
  (drsForDiscourse DRS-3448749395-20705)
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (not (DrsCaseFn DRS-3448749396-20706)))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (possessiveRelation his13541 time13553))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (possessiveRelation time13553 PouryaieVali))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa PouryaieVali Wrestler))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa time13553 TimeInterval))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (properNameReference PouryaieVali))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (willBe (DrsCaseFn DRS-3448749397-20708)))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (toLocation go14856 mosque14889))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa mosque14889 Mosque-Building))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (performedBy see14903 PouryaieVali))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (primaryObjectMoving go14856 PouryaieVali))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa see14903 VisualPerception))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (perceivedThings see14903 (DrsCaseFn DRS-3448749399-20712)))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa go14856 Movement-TranslationEvent))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (infoTransferred say15252
      (and (DrsCaseFn DRS-3448749401-20714)
           (DrsCaseFn DRS-3448749401-20715))))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa say15252 Informing))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (senderOfInfo say15252 PouryaieVali))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (beliefs PouryaieVali (DrsCaseFn DRS-3448749411-20724)))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (beliefs PouryaieVali (DrsCaseFn DRS-3448749418-20731)))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa group-of-option17137 Set-Mathematical))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (implies-DrsDrs (DrsCaseFn DRS-3448749419-20734)
      (DrsCaseFn DRS-3448749419-20735)))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (possessiveRelation PouryaieVali group-of-option17137))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (cardinality group-of-option17137 2))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (isa option17207 ChoiceInSelecting))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (nthInSeries option17207 SERIES17192 1))
    (ist-Information (DrsCaseFn DRS-3448749395-20705)
     (possible (DrsCaseFn DRS-3448749428-20741)))
```

```
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (isa SelectingSomething20744 SelectingSomething))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (implies (chosenItem SelectingSomething20744 option17207)
  (intends PouryaieVali (DrsCaseFn DRS20743))))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (possible-Historical (DrsCaseFn DRS20743)))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (possible (DrsCaseFn DRS-3448749452-20755)))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (nthInSeries option17624 SERIES17598 2))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (isa option17624 ChoiceInSelecting))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (implies (chosenItem SelectingSomething20744 option17624)
  (intends PouryaieVali (DrsCaseFn DRS20758))))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (possible-Historical (DrsCaseFn DRS20758)))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (implies
  (and (intends PouryaieVali make18768)
       (intends PouryaieVali (DrsCaseFn DRS20758)))
  (hasHighMoralValue (DrsCaseFn DRS20758))))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (performedBy SelectingSomething20744 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (age mother16819
  (RelativeGenericValueFn age HumanMother
   highToVeryHighAmountOf)))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (by-Underspecified make18768 SelectingSomething20744))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (causes-SitProp make18768
  (feelsEmotion mother16819
   (MediumToVeryHighAmountFn Happiness))))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (isa make18768 CausingToBeInACertainCondition))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (isa mother16819 HumanMother))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (chosenItem SelectingSomething20744 option17624))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (nthInSeries option17624 SERIES19278 2))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (performedBy make18768 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749395-20705)
 (intends PouryaieVali make18768))
(ist-Information (DrsCaseFn DRS-3448749396-20706)
 (greaterThanOrEqualTo (RenownLevelFn most13459)
  (RenownLevelFn PouryaieVali)))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (isa province14257 Province))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (isa athlete13913 Athlete))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (from-UnderspecifiedLocation athlete13913 province14257))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (age athlete13913
  (RelativeGenericValueFn age (DrsCaseFn DRS-3448749397-20709)
   veryLowToLowAmountOf)))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (isa wrestle13726 WrestlingSportsEvent))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (competingAgents wrestle13726 athlete13913))
(ist-Information (DrsCaseFn DRS-3448749397-20708)
 (competingAgents wrestle13726 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749397-20709)
 (from-UnderspecifiedLocation athlete13913 province14257))
```

```
(ist-Information (DrsCaseFn DRS-3448749397-20709)
 (isa athlete13913 Athlete))
(ist-Information (DrsCaseFn DRS-3448749397-20709)
 (isa province14257 Province))
(ist-Information (DrsCaseFn DRS-3448749399-20712)
 (isa pray15107 Praying))
(ist-Information (DrsCaseFn DRS-3448749399-20712)
 (age athlete13913
  (RelativeGenericValueFn age Athlete veryLowToLowAmountOf)))
(ist-Information (DrsCaseFn DRS-3448749399-20712)
 (mother athlete13913 mother14943))
(ist-Information (DrsCaseFn DRS-3448749399-20712)
 (isa athlete13913 Athlete))
(ist-Information (DrsCaseFn DRS-3448749399-20712)
 (performedBy pray15107 mother14943))
(ist-Information (DrsCaseFn DRS-3448749401-20714)
 (possessiveRelation my15298 son15303))
(ist-Information (DrsCaseFn DRS-3448749401-20714)
 (willBe (DrsCaseFn DRS-3448749401-20716)))
(ist-Information (DrsCaseFn DRS-3448749401-20714)
 (sons my15298 son15303))
(ist-Information (DrsCaseFn DRS-3448749401-20716)
 (isa wrestle15382 WrestlingSportsEvent))
(ist-Information (DrsCaseFn DRS-3448749401-20716)
 (competingAgents wrestle15382 son15303))
(ist-Information (DrsCaseFn DRS-3448749401-20716)
 (competingAgents wrestle15382 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749401-20716)
 (properNameReference PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749401-20715)
 (enables-Generic help15610 (DrsCaseFn DRS-3448749401-20717)))
(ist-Information (DrsCaseFn DRS-3448749401-20715)
 (beneficiary help15610 him15632))
(ist-Information (DrsCaseFn DRS-3448749401-20715)
 (performedBy help15610 (:GAP :SUBJECT)))
(ist-Information (DrsCaseFn DRS-3448749401-20715)
 (isa help15610 Helping-PromotingSomething))
(ist-Information (DrsCaseFn DRS-3448749401-20715)
 (facilitates-SitProp help15610
  (DrsCaseFn DRS-3448749401-20720)))
(ist-Information (DrsCaseFn DRS-3448749401-20717)
 (possible (DrsCaseFn DRS-3448749401-20718)))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (compoundNoun prize15836 money15869))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (purposeInEvent he15756 use15793
  (DrsCaseFn DRS-3448749401-20719)))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (isa money15869 Currency))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (performedBy use15793 he15756))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (isa use15793 UsingAnObject))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (isa prize15836 Prize))
(ist-Information (DrsCaseFn DRS-3448749401-20718)
 (instrument-Generic use15793 money15869))
(ist-Information (DrsCaseFn DRS-3448749401-20719)
 (isa house16022 House-Modern))
(ist-Information (DrsCaseFn DRS-3448749401-20719)
 (isa buy15913 Buying))
(ist-Information (DrsCaseFn DRS-3448749401-20719)
 (objectPaidFor buy15913 house16022))
(ist-Information (DrsCaseFn DRS-3448749401-20719)
 (buyer buy15913 he15756))
(ist-Information (DrsCaseFn DRS-3448749401-20720)
 (isa match15691 MatchSportsEvent))
(ist-Information (DrsCaseFn DRS-3448749401-20720)
```

```
 (winner-First match15691 him15632))
(ist-Information (DrsCaseFn DRS-3448749411-20724)
 (greaterThan (MeasureOfPropFn (DrsCaseFn DRS-3448749411-20725))
  (MeasureOfPropFn (DrsCaseFn DRS-3448749411-20726))))
(ist-Information (DrsCaseFn DRS-3448749411-20724)
 (age athlete13913
  (RelativeGenericValueFn age Athlete veryLowToLowAmountOf)))
(ist-Information (DrsCaseFn DRS-3448749411-20724)
 (isa money15869 Currency))
(ist-Information (DrsCaseFn DRS-3448749411-20724)
 (isa athlete13913 Athlete))
(ist-Information (DrsCaseFn DRS-3448749411-20724)
 (requires-Underspecified athlete13913 money15869))
(ist-Information (DrsCaseFn DRS-3448749411-20725)
 (isa money15869 Currency))
(ist-Information (DrsCaseFn DRS-3448749411-20725)
 (requires-Underspecified athlete13913 money15869))
(ist-Information (DrsCaseFn DRS-3448749411-20726)
 (isa do16423 PurposefulAction))
(ist-Information (DrsCaseFn DRS-3448749411-20726)
 (doneBy do16423 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749418-20731)
 (willBe (DrsCaseFn DRS-3448749418-20732)))
(ist-Information (DrsCaseFn DRS-3448749418-20731)
 (isa match16625 MatchSportsEvent))
(ist-Information (DrsCaseFn DRS-3448749418-20731)
 (winner-First match16625 (:GAP :SUBJECT)))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (age mother16819
  (RelativeGenericValueFn age HumanMother
   highToVeryHighAmountOf)))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (doneBy break16687 win16589))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (isa break16687 BreakingEvent))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (possessiveRelation mother16819 heart16940))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (isa heart16940 Heart-LocusOfFeelings))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (objectOfStateChange break16687 heart16940))
(ist-Information (DrsCaseFn DRS-3448749418-20732)
 (isa mother16819 HumanMother))
(ist-Information (DrsCaseFn DRS-3448749419-20734)
 (member option17137 group-of-option17137))
(ist-Information (DrsCaseFn DRS-3448749419-20735)
 (isa option17137 ChoiceInSelecting))
(ist-Information (DrsCaseFn DRS-3448749428-20741)
 (winner-First match17313 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749428-20741)
 (isa match17313 MatchSportsEvent))
(ist-Information (DrsCaseFn DRS20743)
 (isa keep17408 KeepingSomething))
(ist-Information (DrsCaseFn DRS20743)
 (objectRetained keep17408 status17441))
(ist-Information (DrsCaseFn DRS20743)
 (agentRetaining keep17408 option17207))
(ist-Information (DrsCaseFn DRS20743)
 (isa PouryaieVali Wrestler))
(ist-Information (DrsCaseFn DRS20743) (isa status17441 Reputation))
(ist-Information (DrsCaseFn DRS20743)
 (topicOfIndividual status17441 PouryaieVali))
(ist-Information (DrsCaseFn DRS20743)
 (not (DrsCaseFn DRS-3448749435-20747)))
(ist-Information (DrsCaseFn DRS20743)
 (possessiveRelation PouryaieVali status17441))
(ist-Information (DrsCaseFn DRS-3448749435-20747)
 (greaterThanOrEqualTo
```

```
  (RelativeQualityInGroupFn other17482 Wrestler)
  (RelativeQualityInGroupFn PouryaieVali Wrestler)))
(ist-Information (DrsCaseFn DRS-3448749452-20755)
 (doneBy option17624 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749452-20755)
 (isa match17726 MatchSportsEvent))
(ist-Information (DrsCaseFn DRS-3448749452-20755)
 (purposeInEvent PouryaieVali option17624
  (DrsCaseFn DRS-3448749452-20756)))
(ist-Information (DrsCaseFn DRS-3448749452-20755)
 (isa option17624 LosingAConflict))
(ist-Information (DrsCaseFn DRS-3448749452-20755)
 (topicOfIndividual option17624 match17726))
(ist-Information (DrsCaseFn DRS-3448749452-20756)
 (causes-SitProp make17769
  (feelsEmotion mother16819
   (MediumToVeryHighAmountFn Happiness))))
(ist-Information (DrsCaseFn DRS-3448749452-20756)
 (isa mother16819 HumanMother))
(ist-Information (DrsCaseFn DRS-3448749452-20756)
 (age mother16819
  (RelativeGenericValueFn age HumanMother
   highToVeryHighAmountOf)))
(ist-Information (DrsCaseFn DRS-3448749452-20756)
 (performedBy make17769 PouryaieVali))
(ist-Information (DrsCaseFn DRS-3448749452-20756)
 (isa make17769 CausingToBeInACertainCondition))
(ist-Information (DrsCaseFn DRS20758)
 (objectActedOn risk18124 status18157))
(ist-Information (DrsCaseFn DRS20758) (isa risk18124 RiskTaking))
(ist-Information (DrsCaseFn DRS20758)
 (possessiveRelation PouryaieVali status18157))
(ist-Information (DrsCaseFn DRS20758)
 (performedBy risk18124 option17624))
(ist-Information (DrsCaseFn DRS20758)
 (facilitates-EventEvent option17624
  (DrsCaseFn DRS-3448749458-20761)))
(ist-Information (DrsCaseFn DRS20758) (isa status18157 Reputation))
(ist-Information (DrsCaseFn DRS-3448749458-20761)
 (isa athlete13913 Athlete))
(ist-Information (DrsCaseFn DRS-3448749458-20761)
 (isa buy18306 Buying))
(ist-Information (DrsCaseFn DRS-3448749458-20761)
 (objectPaidFor buy18306 house18508))
(ist-Information (DrsCaseFn DRS-3448749458-20761)
 (buyer buy18306 athlete13913))
(ist-Information (DrsCaseFn DRS-3448749458-20761)
 (isa house18508 House-Modern))
(ist-Information (DrsCaseFn DRS-3448749458-20761)
 (age athlete13913
      (RelativeGenericValueFn age Athlete veryLowToLowAmountOf)))))
```

8.2. Appendix B: Rules

The rules described in this section are back-chaining rules. These rules, unlike Prolog rules are unordered and return all answers that satisfy the constraints mentioned in them. However, they are arranged so that exactly one of the possible decision modes returns an answer.

Most of the predicates used are from the Research Cyc KB. For details about these predicates please refer to the Research Cyc ontology. Before describing the rules, I will briefly describe the special predicates and list those defined for the purpose of this thesis.

8.2.1. Special Predicates and Predicates Added to the KB

`Wmexclusive`**:** is a wrapper that binds :facts to :`wmExclusively`. It assumes all structural `isa`'s and `genl`'s are in the working memory and not in the KB nor the structural cache. This predicate was used in Fire v2, but now works similar to `Wmonly`.

`uninferredSentence`: is a meta-knowledge predicate. It is similar to `not` except that the statements are justified by timestamps.

`drsForDiscourse`: links a DRS id to a discourse id.

`DrsCaseFn`: takes a DRS id and returns a WM case that holds the facts in that DRS.

`SubexpressionMatching`: means some pattern unifies with some subexpression of an expression.

The following are predicates added to the KB for reasoning about moral scenarios:

```
(isa preventedSit Predicate)
(arity preventedSit 2)
```

```
(argisa preventedSit 1 Thing)
(argisa preventedSit 2 Event)
(comment preventedSit
 “(preventedSit ?Thing1 ?Event) indicates that Thing1 prevented Thing2 from happening”)

(isa sacredValue Predicate)
(arity sacredValue 2)
(argisa sacredValue 1 Collection)
(argisa sacredValue 2 Event)
(comment sacredValue
 “(sacredValue ?Collection ?Event) indicates that in the event of ?Event ?Collection are
     considered a sacred value”)

(isa hasNegativeUtil Predicate)
(arity hasNegativeUtil 1)
(argisa hasNegativeUtil 1 Event)
(comment hasNegativeUtil
 “(hasNegativeUtil ?Event) is use to indicate that ?Event happening has a negative utility”)

(isa hasHighMoralValue Predicate)
(arity hasHighMoralValue 1)
(argisa hasHighMoralValue 1 Event)
(comment hasNegativeUtil
 “(hasHighMoralValue ?Event) indicates that that performing ?Event has high moral value”)

(isa consideredMoralAct Predicate)
(arity consideredMoralAct 2)
(argisa consideredMoralAct 1 Thing)
(argisa consideredMoralAct 2 Event)
(comment consideredMoralAct
 “(consideredMoralAct ?Thing ?Event) indicates that that performing ?Thing at ?Event is
     considered a moral value”)

(isa makeDecision Predicate)
(arity makeDecision 1)
(argisa makeDecision 1 Event)
(comment makeDecision
 “(makeDecision ?Event) indicates the choice that was made in a decision scenario”)

(isa ordersOfMagnitudeDifferent Predicate)
(arity ordersOfMagnitudeDifferent 1)
(argisa ordersOfMagnitudeDifferent 1 Event)
(comment ordersOfMagnitudeDifferent
“(ordersOfMagnitudeDifferent ?Event) indicates that there is an order of magnitude
     difference between the utility of choices involved in the decision”)
```

```
(isa notOrdersOfMagnitudeDifferent Predicate)
(arity notOrdersOfMagnitudeDifferent 1)
(argisa notOrdersOfMagnitudeDifferent 1 Event)
(comment ordersOfMagnitudeDifferent
“(notOrdersOfMagnitudeDifferent ?Event) indicates that there is not an order of magnitude
     difference between the utility of choices involved in the decision”)

(isa involvesSacredValue Predicate)
(arity involvesSacredValue 1)
(argisa involvesSacredValue 1 Event)
(comment pureDeontologicalChoice
“(pureDeontologicalChoice ?Event) ?Event here is the decision to be made, and the predicate
     indicates that the pure-deontological type of reasoning was used for that decision”)

(isa utilDeontologicalChoice Predicate)
(arity utilDeontologicalChoice 1)
(argisa utilDeontologicalChoice 1 Event)
(comment utilDeontologicalChoice
“(utilDeontologicalChoice ?Event) ?Event here is the decision to be made, and the predicate
     indicates that the utilitarian-deontological type of reasoning was used for that
     decision”)


(isa deontologicalChoice Predicate)
(arity deontologicalChoice 1)
(argisa deontologicalChoice 1 Event)
(comment deontologicalChoice
“(deontologicalChoice ?Event) ?Event here is the decision to be made, and the predicate
     indicates that the deontological type of reasoning was used for that decision”)

(isa utilitarianChoice Predicate)
(arity utilitarianChoice 1)
(argisa utilitarianChoice 1 Event)
(comment utilitarianChoice
“(utilitarianChoice?Event) ?Event here is the decision to be made, and the predicate
     indicates that the utilitarian type of reasoning was used for that decision”)

(isa highestUtilChoice Predicate)
(arity highestUtilChoice 1)
(argisa highestUtilChoice 1 Event)
(comment highestUtilChoice
“(highestUtilChoice ?Event) indicates whether ?Event, which is the choice, is the choice with
     the highest utility”)

(isa isInaction Predicate)
(arity isInaction 1)
```

```
(argisa isInaction 1 Event)
(comment isInaction “(isInaction ?Event) indicates that an event is an inaction)

(isa ActivatingPoweredDevice Collection)
(isa OperateInNatureReserve Event-Localized)
(genls ChoiceInSelecting SelectingSomething)
(genls Railway-siding Railway)
(genls Starvation Dying)
(genls Extinction Dying)
```

These are events defined in the KB which are known to have negative utilities.

```
(hasNegativeUtil Dying)
(hasNegativeUtil Deforestation)
(hasNegativeUtil Extinction)
(hasNegativeUtil Starvation)
(hasNegativeUtil EmployeeLayoff)
(hasNegativeUtil HarmingAnAgent)
```

The following are the sacred values defined in the KB.

```
(sacredValue HumanChild Dying)
(sacredValue AnimalSpecies Extinction)
(sacredValue Fish Extinction)
(sacredValue Person Starvation)
(sacredValue Person Dying)
(sacredValue MedicalPatient Dying)
(sacredValue Forest Deforestation)
(sacredValue FactoryWorker EmployeeLayoff)
(sacredValue HomelessPerson HarmingAnAgent)
(sacredValue Army-BranchOfService OperateInNatureReserve)
```

8.2.2. High-Level FPR Rules

The highest level rule called by MoralDM is `makeDecision`.

```
(<== (makeDecision ?choice)
     (utilCalculation)
     (decisionMaker ?choice))
```

This rule in turn calls `utilCalculation` and `decisionMaker`. For more information about `utilCalculation` please look at the next section. This rule returns the result of the OMR module which is the qualitative relationship between the utility of the choices. decisionMaker fires two sets of rules: utilitarianChoice and deontologicalChoice.

```
(<== (decisionMaker ?choice)
     (deontologicalChoice ?choice))

(<== (decisionMaker ?choice)
     (utilitarianChoice ?choice))
```

These methods are mutually exclusive, returning at most one choice per scenario. deontologicalChoice in turn calls pureDeontologicalChoice and utilDeontologicalChoice.

```
(<== (deontologicalChoice ?choice)
     (pureDeontologicalChoice ?choice))

(<== (deontologicalChoice ?choice)
     (utilDeontologicalChoice ?choice))

(<== (pureDeontologicalChoice ?choice)
     (involvesSacredValue ?choice)
     (choices ?decision ?choice)
     (notOrdersOfMagnitudeDifferent ?decision)
     (isInaction ?choice))

(<== (utilDeontologicalChoice ?choice)
     (involvesSacredValue ?choice)
     (choices ?decision ?choice)
     (ordersOfMagnitudeDifferent ?decision)
     (highestUtilChoice ?choice))

(<== (utilitarianChoice ?choice)
     (choicesAre ?choice)
     (uninferredSentence
      (involvesSacredValue ?choice))
     (highestUtilChoice ?choice))
```

If there are no protected values involved in the scenario deontologicalChoice fails and utilitarianChoice returns the choice with the highest utility. If there are sacred values involved, then utilitarianChoice fails and deontologicalChoice return the results of one of these subsequent rules: utilDeontologicalChoice and pureDeontologicalChoice.

Again these two methods are mutually exclusive and only one can come up with an answer for any given scenario. If there is an order of magnitude difference between the utility of choices, `utilDeontologicalChoice` returns the option with highest utility. Otherwise, the choice which does not violate the psychological findings on moral decision making discussed previously is returned by `pureDeontologicalChoice`.

The following rules are the rules used for determining whether or not a sacred value exists in the scenario:

```
(<== (involvesSacredValue ?choice)
     (choicesAndConsequences ?choice ?consequence)
     (typeOfGroupEffected ?consequence ?groupEffected)
     (SacredValue ?groupEffected ?sacredAction)
     (isa ?consequence ?sacredAction))


(<== (involvesSacredValue ?choice)
     (causes-PropSit (chosenItem ?select ?choice) ?consequence)
     (wmExclusively (isa ?consequence ?typeOfConcesequnce))
     (objectOfStateChange ?consequence ?y)
     (wmExclusively (isa ?y ?typeOfY))
     (SacredValue ?typeOfConcesequnce ?typeOfY))



(<== (involvesSacredValue ?choice)
     (causes-PropSit (chosenItem ?select ?choice) ?consequence)
     (isa ?consequence ?typeOfConcesequnce)
     (relationInstanceMember objectOfStateChange ?consequence ?y)
     (relationMemberInstance isa ?y ?typeOfY)
     (SacredValue ?typeOfY ?typeOfConcesequnce))


(<== (involvesSacredValue ?choice)
     (choicesAndConsequences ?choice ?consequence)
     (preventedProp ?consequence ?typeOfConsequence)
     (toPossessor ?typeOfConsequence ?x)
     (objectOfPossessionTransfer ?typeOfConsequence ?y)
     (wmExclusively (isa ?x ?typeOfX))
     (wmExclusively (isa ?y ?typeOfY))
     (SacredValue ?pred (?typeOfX ?typeOfY))
     (isa ?typeOfConsequence ?pred))


(<== (involvesSacredValue ?choice)
     (causes-SitProp ?choice ?conjunction)
     (subexpressionMatching (objectOfStateChange ?consequence ?area)
                            ?conjunction
                            ?match-expr2)
     (isa ?consequence ?typeOfConcesequnce)
     (isa ?area ?typeOfY)
     (SacredValue ?typeOfConcesequnce ?typeOfY))
```

In each case, the goal is to find the consequence of each action, and for each consequence (direct and indirect) determine whether or not a known sacred value is affected. In some cases this requires investigating the types, and membership relation of each thing. Then determining whether those parent types and parent members are sacred or not.

Sometimes it is necessary to check the type of group affected:

```
(<== (typeOfGroupEffected ?consequence ?typeOf)
     (objectActedOn ?consequence ?groupEffected)
     (isa ?groupEffected (GroupFn ?typeOf)))

(<== (typeOfGroupEffected ?consequence ?typeOf)
     (objectActedOn ?consequence ?groupEffected)
     (generalizes ?groupEffected ?typeOf))

(<== (typeOfGroupEffected ?consequence ?typeOf)
     (beneficiary ?consequence ?benefic)
     (relationMemberInstance isa ?benefic ?typeOf))

(<== (typeOfGroupEffected ?consequence ?typeOf)
     (organismKilled ?consequence ?killed)
     (isa ?killed ?typeOf))
```

The following rules are used to find the choices, and the consequences of each choice.

```
(<== (choicesAndConsequences ?choices ?consequences)
     (choicesAre ?choices)
     (causes-PropProp
      (chosenItem ?selecting ?choices) ?consequences))


(<== (choicesAndConsequences ?choices ?secondConsq)
     (choicesAre ?choices)
     (causes-PropSit
      (chosenItem ?selecting ?choices) ?consequences)
     (by-Underspecified ?consequences ?x)
     (causes-SitSit ?x ?secondConsq))


(<== (choicesAre ?choices)
     (drsForDiscourse ?x)
     (ist-Information (DrsCaseFn ?x)
       (isa ?choices ChoiceInSelecting))
     (uninferredSentence  (ist-Information (DrsCaseFn ?y)
                            (member ?choices ?z))))
```

Sometimes it is required to know the type of consequence. The following rules are used to distinguish between consequences of prevention vs. promotion.

```
(<== (consequenceOfPrevention ?choice ?consequence)
     (causes-PropProp
      (chosenItem ?selecting ?choice) ?consequence)
     (isa ?consequence PreventingSomething))


(<== (consequenceOfPrevention ?choice ?type)
     (causes-PropSit
      (chosenItem ?selecting ?choice) ?consequence)
     (causes-SitSit ?consequence ?type)
     (isa ?type PreventingSomething))


(<== (consequenceOfPromotion ?choice ?consequence)
     (causes-PropProp
      (chosenItem ?selecting ?choice) ?consequence)
     (uninferredSentence (isa ?consequence PreventingSomething)))



(<== (consequenceOfPromotion ?choice ?consequence)
     (causes-PropSit
      (chosenItem ?selecting ?choice) ?consequence)
     (uninferredSentence (isa ?consequence PreventingSomething))
     (uninferredSentence (causes-SitSit ?consequence ?type)))


(<== (consequenceOfAction ?choice ?p)
     (choicesAndConsequences ?choice ?p)
     (doneBy ?kill ?choice)
     (organismKilled ?kill ?killed)
     (uninferredSentence (isa ?p RescuingSomeone)))
```

The following rules are used to determine the patient of harm:

```
(<== (patientEffectedByAction ?choice ?patient)
     (choicesAndConsequences ?choice ?p)
     (causes-SitProp ?choice ?conjunction)
     (subexpressionMatching (hasPhysiologicalFeature ?patient Paraplegia)
                            ?conjunction
                            ?match-expr)))

(<== (patientEffectedByAction ?choice ?member)
     (choicesAndConsequences ?choice ?p)
     (doneBy ?action ?choice)
     (primaryObjectMoving ?action ?object)
     (relationInstanceMember with-UnderspecifiedAgent ?object ?member))

(<== (patientEffectedByAction ?choice ?set-of-patients)
     (choicesAndConsequences ?choice ?p)
     (causes-SitProp ?choice ?conjunction)
     (subexpressionMatching (objectActedOn ?cause ?patient)
                            ?conjunction
                            ?match-expr)
     (relationInstanceMember with-UnderspecifiedAgent ?patient ?set-of-patients))
```

In some scenarios MoralDM had to be able to distinguish between the choice which brings benefits to the agent itself from a choice which brings utility to another agent. The following rules are used for this task:

```
(<== (choiceBenefitingSelf ?choice)
     (drsForDiscourse ?x)
     (ist-Information (DrsCaseFn ?x)
       (implies (chosenItem ?selectingSomething ?choice)
                (intends ?agent (DrsCaseFn ?something))))
     (ist-Information (DrsCaseFn ?something)
       (possessiveRelation ?agent ?agentPossess))
     (ist-Information (DrsCaseFn ?something)
     (performedBy ?agentsAction ?choice))
     (ist-Information (DrsCaseFn ?something)
     (beneficiary ?agentsAction ?agentPossess))


(<== (choiceBenefitingSelf ?choice)
     (drsForDiscourse ?x)
     (ist-Information (DrsCaseFn ?x)
       (implies (chosenItem ?selectingSomething ?choice)
                (intends ?agent (DrsCaseFn ?something))))
     (ist-Information (DrsCaseFn ?something)
       (possessiveRelation ?agent ?agentPossess))
     (ist-Information (DrsCaseFn ?something)
       (isa ?agentPossess Reputation))
     (ist-Information (DrsCaseFn ?something)
       (agentRetaining ?keep ?choice)
     (ist-Information (DrsCaseFn ?something)
       (objectRetained ?keep ?agentPossess)))
```

The following rule was used to determine whether or not the choice involved the agent sacrificing his life.

```
(<== (involvesSacrificingOwnLife ?choice ?something)
     (ist-Information (DrsCaseFn ?something)
       (possessiveRelation ?agent ?agentPossess))
     (ist-Information (DrsCaseFn ?something)
       (isa ?agentPossess Living))
     (ist-Information (DrsCaseFn ?something)
       (isa ?sac Sacrifice))
     (ist-Information (DrsCaseFn ?something)
       (objectActedOn ?sac ?agentPossess))
     (ist-Information (DrsCaseFn ?something)
       (performedBy ?sac ?choice)))
```

8.2.3. High-Level Rules for Calculating the Utility of Choices

In this section some of the high-level rules for calculating the utility of choices are described. Each choice has a positive side effect and a negative one. For example, going back to the starvation scenario, the choice of rerouting will save 1000 people and will kill 100 people. Therefore, the system needs to figure not only the out the cardinality of each of these outcomes, it has to figure out which ones are positive and which ones are negative. Moreover, the system needs to find out what amount of quantity sensitivity, *k*, it should use for calculating the order of magnitude relation between the utilities. For that, what needs to be determined is whether or not the case includes a sacred value, and if so what the causal structure of it is.

```
(<== (utilCalculation)
     (utilOfChoices ?utils)
     (compareUtils ?utils))

(<== (utilOfChoices ?utils)
     (choicesAre ?choice)
     (utilOfChoice ?choice ?util)
     (evaluate ?utils
               (TheClosedRetrievalSetOf
                (?choice2 ?util2)
                (utilOfChoice ?choice2 ?util2))))
```

These are some of the rules used to determine the utility of each choice. For a complete list of them please refer to rules.lsp. Each of these rules determines first whether the consequence of the choice is that of a promotion or of a prevention. Then it calls to determine `isNegativeUtil` whether it has negative utility. Next, `cardinalityOfThingsEffectedByPromotion` or `cardinalityOfThingsEffectedByPrevention` are called. Next, the initial *k* is calculated as described in Chapter 3. Having the initial value of *k*, the system can now determine ε. Using *k* and ε, the order of magnitude between the utilities are calculated by calling `utilRelation`.

```
(<== (utilOfChoice ?choice ?util)
     (consequenceOfPromotion ?choice ?consequence)
     (isNegativeUtil ?consequence)
     (cardinalityOfThingsEffectedByPromotion ?choice ?promotionCardinality)
     (evaluate ?promotionUtil (TimesFn ?promotionCardinality -1))
     (cardinalityOfThingsEffectedByPrevention ?choice ?preventionCardinality)
     (findInitialK ?promotionUtil ?preventionCardinality ?k)
     (findE ?k ?e)
     (determineK ?choice ?k ?e ?z ?k-result)
     (utilRelation ?promotionUtil ?preventionCardinality ?k-result ?util))
```

```
(<== (utilOfChoice ?choice ?util)
     (consequenceOfPrevention ?choice ?consequence)
     (isNegativeUtil ?consequence)
     (cardinalityOfConsequencesEffectedByPrevention ?consequence ?preventionCardinality)
     (evaluate ?preventionUtil (TimesFn ?preventionCardinality -1))
     (consequenceOfPrevention ?choice ?secondConsequence)
     (uninferredSentence (isNegativeUtil ?secondConsequence))
     (cardinalityOfConsequencesEffectedByPrevention ?secondConsequence
             ?secondPreventionCardinality)
     (findInitialK ?preventionUtil ?secondPreventionCardinality  ?k)
     (findE ?k ?e)
     (determineK ?choice ?k ?e ?z ?k-result)
     (utilRelation ?preventionUtil ?secondPreventionCardinality ?k-result ?util))
```

```
(<== (utilOfChoice ?choice ?util)
     (causes-PropSit (chosenItem ?selecting ?choice) ?consequence)
     (causes-SitSit ?consequence ?preventedCause)
     (preventedProp ?preventedCause ?conjunctionPrevented)

     (subexpressionMatching  (relationMemberInstance hasAttributes ?whom ?what)
                            ?conjunctionPrevented
                            ?match-expr)

     (subexpressionMatching (cardinality ?whom ?preventionUtil)
                            ?conjunctionPrevented
                            ?match-exprr)

     (causes-SitProp ?choice ?conjunction)
     (subexpressionMatching  (relationMemberInstance hasAttributes ?worker ?x)
                            ?conjunction
                            ?match-exprrr)

     (subexpressionMatching (cardinality ?worker ?numberUnemployed)
                            ?conjunction
                            ?match-exprrrr)


     (evaluate ?promotionCardinality (TimesFn ?numberUnemployed -1))
     (findInitialK ?preventionUtil ?promotionCardinality ?k)
     (findE ?k ?e)
     (determineK ?choice ?k ?e ?z ?k-result)
     (utilRelation ?preventionUtil ?promotionCardinality ?k-result ?util))
```

The following are used to calculate the value of *k*. First, they try to figure out whether a sacred value exists in the scenario or not. Then they call `determineKRule` which determines the type of sensitivity which should be used given the causal structure of the scenario. Finally, the initial value of *k*, ε, the information about the existence of sacred values in the scenario and the causal structure are used to calculate the appropriate *k* for the scenario.

```
(<== (determineK ?choice ?k ?e ?z ?k-result)
     (involvesSacredValue ?choice)
     (determineKRule ?choice ?k ?e ?z)
```

```
    (evaluate ?k-result (PlusFn ?z ?e))

(<== (determineK ?choice ?k ?e ?z ?k-result)
     (uninferredSentence
      (involvesSacredValue ?choice))
     (determineKRule ?choice ?k ?e ?z)
     (evaluate ?k-result (PlusFn ?z (TimesFn ?e -1)))
```

These rules are used to determine what degree of quantity sensitivity should be used by determining the causal structure of the scenario.

```
(<== (determineKRule ?choice ?k ?e ?z)
     (withLowQuantitySensitivity ?choice ?k ?e ?z))

(<== (determineKRule ?choice ?k ?e ?z)
     (withHighQuantitySensitivity ?choice ?k ?e ?z))

(<== (determineKRule ?choice ?k ?e ?z)
     (withDefaultQuantitySensitivity ?choice ?k ?e ?z))

(<== (withLowQuantitySensitivity ?choice ?k ?e ?z)
      (lessQuantitySensitive ?choice)
     (uninferredSentence
      (moreQuantitySensitive ?choice))
     (evaluate ?z (PlusFn ?k ?e)))

(<== (withHighQuantitySensitivity ?choice ?k ?e ?z)
      (moreQuantitySensitive ?choice)
     (uninferredSentence
      (lessQuantitySensitive ?choice))
     (evaluate ?z (MinusFn ?k ?e)))

(<== (withDefaultQuantitySensitivity ?choice ?k ?e ?z)
     (uninferredSentence
      (moreQuantitySensitive ?choice))
     (uninferredSentence
      (lessQuantitySensitive ?choice))
     (evaluate ?z (PlusFn ?k 0)))

(<== (moreQuantitySensitive ?choice)
     (doubleEffect ?choice))

(<== (lessQuantitySensitive ?choice)
     (patientIntervention ?choice))
```

Given Waldmann and Dieterich's (2007) findings, patient intervention decreases the quantity sensitivity and double-effect increases it.

```
(<== (patientIntervention ?choice)
     (causes-PropSit
```

```
 (chosenItem ?selecting ?choice) ?consequence)
(typeOfGroupEffected ?consequence ?typeOf)
(objectActedOn ?choice ?typeOfChoice)
(isa ?typeOfChoice ?typeOf))

(<==(doubleEffect ?choice)
    (choicesAndConsequences ?choice ?directConsequence)
    (causes-PropProp (chosenItem ?selecting ?directConsequence) ?indirectConsequence)
    (isConsequenceSacredValue ?indirectConsequence))
```

8.2.4. High-Level Rules for Reasoning about Candidate Inferences

As discussed in Chapter 3, a previously solved case has to satisfy a number of "if-then rules" (Weber, Ames and Blais 2005) before it can be used as a valid analog for making a decision in the new case. These constraints are described in chapter 3. Here I describe the rules for the third constraint: rules for figuring out whether the inferred intention of the target matches that of the base.

These rules deal with candidate inferences returned by SME. That is, once SME forms a mapping and the threshold of the match is higher than the threshold, candidate inferences are added to the working memory. The base and the target cases need to either both have moral values, or both not be in the sacred domain. The first set of rules is fired to determine whether the inferred intention of the agent has a moral value or not. These rules in turn calls `moralityToBeEvaluated` which determines what the moral value in question is and whether it is described in the KB to have a moral value or not. Once the system finds the `DrsCaseFn` in which the moral value is defined it checks for certain other constraints. For example, it checks whether or not that particular choice will hurt the agent making the choice or will it hurt another agent. Also, it checks whether or not that choice has an achievable outcome, and whether the consequence of that choice hurts another fellow Muslim or not. If the inferred decision passes these constraints it is accepted as a valid analog.

```
(<== (hasMoralValue ?choice)
     (moralityToBeEvaluated ?y ?choice)
     (ist-Information (DrsCaseFn ?y) (isa ?life Living))
     (ist-Information (DrsCaseFn ?y) (possessiveRelation ?agent ?life))
     (ist-Information (DrsCaseFn ?whatever) (isa ?agent ?something))
     (genls ?something Person)
     (ist-Information (DrsCaseFn ?y) (isa ?risk RiskTaking))
     (ist-Information (DrsCaseFn ?y) (performedBy ?risk ?choice))
     (achievableOutcome ?choice)
     (uninferredSentence
      (hurtsMuslims ?choice)))
```

```
(<== (hasMoralValue ?choice)
     (moralityToBeEvaluated ?y ?choice)
     (ist-Information (DrsCaseFn ?y) (isa ?life Living))
     (ist-Information (DrsCaseFn ?y) (possessiveRelation ?agent ?life))
     (ist-Information (DrsCaseFn ?whatever) (isa ?agent ?something))
     (genls ?something Person)
     (ist-Information (DrsCaseFn ?y) (isa ?sac Sacrifice))
     (ist-Information (DrsCaseFn ?y) (performedBy ?sac ?choice))
     (ist-Information (DrsCaseFn ?y) (objectActedOn ?sac ?life))
     (achievableOutcome ?choice)
     (uninferredSentence (hurtsMuslims ?choice)))
```

```
(<== (hasMoralValue ?choice)
     (moralityToBeEvaluated ?y ?choice)
     (ist-Information (DrsCaseFn ?y) (performedBy ?help ?choice))
     (ist-Information (DrsCaseFn ?y) (isa ?help Helping-PromotingSomething))
     (ist-Information (DrsCaseFn ?y) (beneficiary ?help ?someone))
     (ist-Information (DrsCaseFn ?y) (facilitates-SitProp ?help (DrsCaseFn ?something)))
     (ist-Information (DrsCaseFn ?something)
                      (age ?someone
                            (RelativeGenericValueFn age ?boo veryLowToLowAmountOf)))
     (ist-Information (DrsCaseFn ?something) (eventHonors ?forWhat ?someone))
     (achievableOutcome ?choice)
     (uninferredSentence (hurtsMuslims ?choice)))
```

This is the rule finds the `DrsCaseFn` in which the moral value is defined.

```
(<== (moralityToBeEvaluated ?y ?choice)
     (intentionIs ?choice ?intentionOfTheAgent)
       (ist-Information (DrsCaseFn ?z)
         (implies (and (intends ?agent ?firstIntention)
                       (intends ?agent (DrsCaseFn ?intentionOfTheAgent)))
                  (hasHighMoralValue (DrsCaseFn ?y))))
```

The `intentionIs` rule determines what the intention of the agent is for choosing that particular choice.

```
(<== (intentionIs ?choice ?intentionOfTheAgent)
     (drsForDiscourse ?x)
```

```
(ist-Information (DrsCaseFn ?x)
  (implies (chosenItem ?selection ?choice)
           (intends ?agent (DrsCaseFn ?intentionOfTheAgent))))
```

This is the rule determines if one of the outcomes is less probably to achieve than the other one or not.

```
(<== (achievableOutcome ?choice)
     (drsForDiscourse ?x)
     (uninferredSentence
      (ist-Information (DrsCaseFn ?x)
       (lessProbableToAchieveOutcome ?selectingSomething ?choice ?choiceB)))
```

The `hurtsMuslims` rule is used to see if the outcome of the choice hurts someone who is a Muslim.

```
(<== (hurtsMuslims ?choice)
     (drsForDiscourse ?y)
     (ist-Information (DrsCaseFn ?y) (subEvents  ?choice ?subevent))
     (ist-Information (DrsCaseFn ?x) (purposeInEvent ?someone ?subevent
                                          (DrsCaseFn ?something))
     (ist-Information (DrsCaseFn ?something)
                      (causes-ThingProp ?someone
                            (blockedPath ?stop ?objectStoped)))
     (ist-Information (DrsCaseFn ?z) (performedBy ?help ?objectStoped))
     (ist-Information (DrsCaseFn ?z) (isa ?help Helping-PromotingSomething))
     (ist-Information (DrsCaseFn ?z) (beneficiary ?help ?helped))
     (hasBeliefSystems ?helped Islam))
```

8.3. Appendix C: A Sample Worked Solution

The following is the worked solution for the starvation scenario described in chapter 3. The representation for this scenario and the closest analogy of it are presented in Appendix A. The first section shows the output of the FIRE reasoning engine when it's using the rules in the FPR module to solve the case. The second section shows the output of SME when matching the starvation scenario with the transfer-of-funds scenario.

8.3.1. FPR Module's Reasoning Trace

The reasoning trace for solving the starvation scenario is online and available in two different formats. The graphML version of this file can be downloaded from:

www.cs.northwestern.edu/~mde345/Thesis/starvation.xml

The text output FIRE Profiler can be downloaded from:

www.cs.northwestern.edu/~mde345/Thesis/starvation.txt

8.3.2. SME Mapping and Candidate Inferences

The following screenshot show the mapping between the starvation and the transfer-of-funds scenarios calculated by SME. The normalized similarity score between the two cases was determined to be 0.474.

Mapping 88

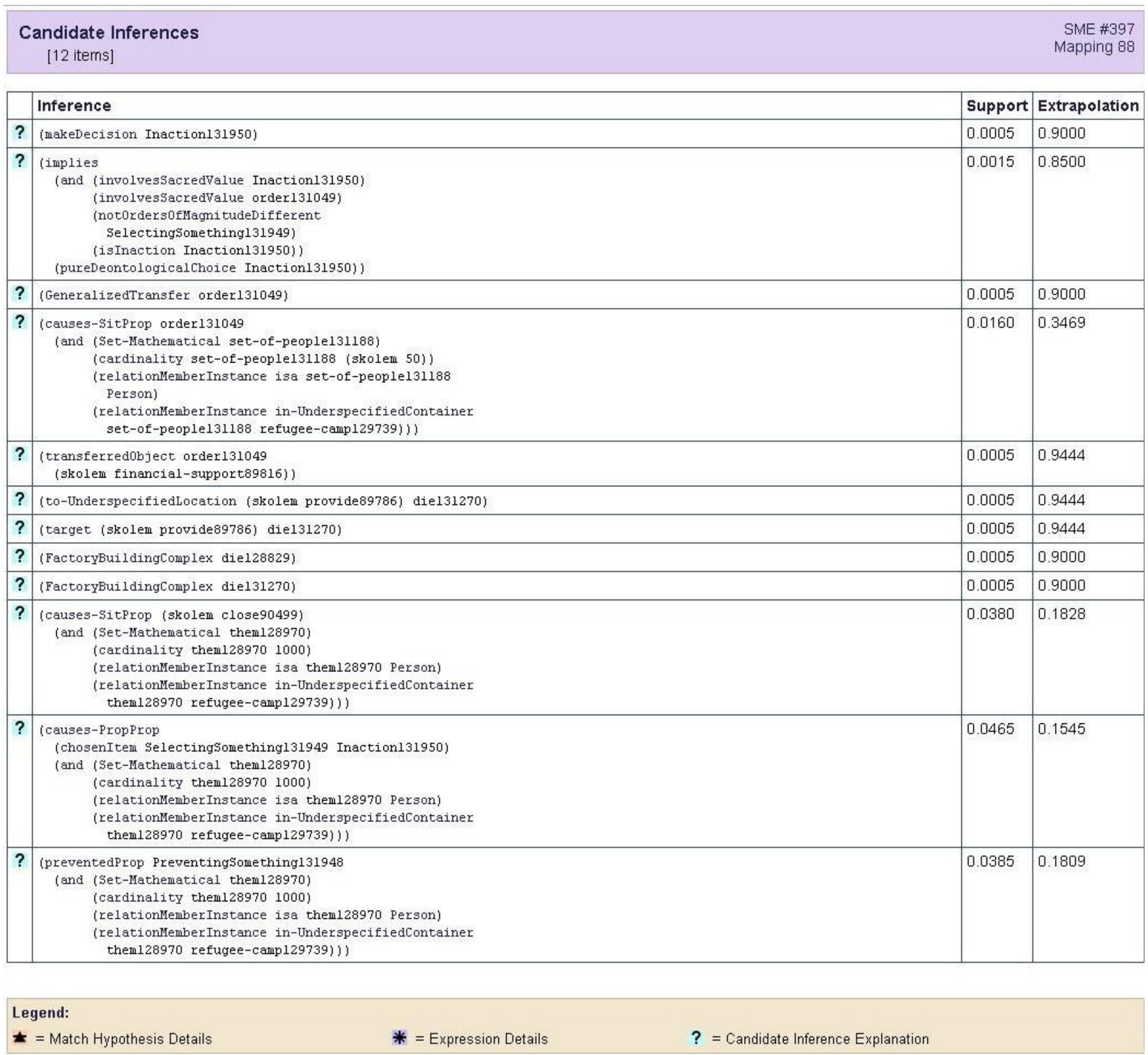
Score: 0.3750
Base: Case0
Target: Case1

Support	Base Item	Target Item
(6)	save90838	save128937
(6)	set-of-worker90320	them128970
(5)	SelectingSomething91623	SelectingSomething131949
(4)	transfer91105	order131049
(3)	Inaction91624	Inaction131950
(3)	you90799	you128898
(3)	PreventingSomething91622	PreventingSomething131948
(2)	set-of-worker90031	set-of-people131188
(2)	hasWorkers	objectOfStateChange
(2)	UnemployedPerson	Person
(2)	hasAttributes	isa
(1)	plant89874	die131270
(1)	plant90251	die128829
(1)	isa	in-UnderspecifiedContainer
(1)	Employee	refugee-camp129739
(1)	500	1000

- 12 candidate inferences
- 22 expression correspondences.
- 15 functor correspondences.

Legend:
= Match Hypothesis Details = Expression Details ? = Candidate Inference Explanatio

Figure 22: Screenshot of the SME mapping between the starvation and the transfer-of-funds scenarios



Candidate Inferences
[12 items]

SME #397
Mapping 88

	Inference	Support	Extrapolation
?	(makeDecision Inaction131950)	0.0005	0.9000
?	(implies (and (involvesSacredValue Inaction131950) (involvesSacredValue order131049) (notOrdersOfMagnitudeDifferent SelectingSomething131949) (isInaction Inaction131950)) (pureDeontologicalChoice Inaction131950))	0.0015	0.8500
?	(GeneralizedTransfer order131049)	0.0005	0.9000
?	(causes-SitProp order131049 (and (Set-Mathematical set-of-people131188) (cardinality set-of-people131188 (skolem 50)) (relationMemberInstance isa set-of-people131188 Person) (relationMemberInstance in-UnderspecifiedContainer set-of-people131188 refugee-camp129739)))	0.0160	0.3469
?	(transferredObject order131049 (skolem financial-support89816))	0.0005	0.9444
?	(to-UnderspecifiedLocation (skolem provide89786) die131270)	0.0005	0.9444
?	(target (skolem provide89786) die131270)	0.0005	0.9444
?	(FactoryBuildingComplex die128829)	0.0005	0.9000
?	(FactoryBuildingComplex die131270)	0.0005	0.9000
?	(causes-SitProp (skolem close90499) (and (Set-Mathematical them128970) (cardinality them128970 1000) (relationMemberInstance isa them128970 Person) (relationMemberInstance in-UnderspecifiedContainer them128970 refugee-camp129739)))	0.0380	0.1828
?	(causes-PropProp (chosenItem SelectingSomething131949 Inaction131950) (and (Set-Mathematical them128970) (cardinality them128970 1000) (relationMemberInstance isa them128970 Person) (relationMemberInstance in-UnderspecifiedContainer them128970 refugee-camp129739)))	0.0465	0.1545
?	(preventedProp PreventingSomething131948 (and (Set-Mathematical them128970) (cardinality them128970 1000) (relationMemberInstance isa them128970 Person) (relationMemberInstance in-UnderspecifiedContainer them128970 refugee-camp129739)))	0.0385	0.1809

Legend:
★ = Match Hypothesis Details ✱ = Expression Details ? = Candidate Inference Explanation

Figure 23: Screenshot of candidate inferences from the transfer-of-funds scenario to the starvation scenario calculated by SME

8.4. Appendix D: Stimuli used in Chapter 4

8.4.1. Farsi version of the stimuli

8.4.1.1. Pourya Vali

Surface change (ΔSF):

علي معروفترين بازيكن پينگ‌پونگ در شهر خود بود. صبح قبل از يك مسابقه با يك بازيكن جوان از شهری ديگر، علي در حالي‌كه در خارج از استاديوم قدم مي‌زد، به پيرزني برخورد كرد كه اينگونه برای فرزند خود دعا می كرد « خدايا، پسرم امروز قرار است با علي مسابقه بدهد. خدايا به او كمک كن تا بتواند در اين مسابقه برنده شود تا بتواند با جايزه مسابقه ازدواج نمايد». علي دو انتخاب داشت مسابقه را ببرد و موقعيت خود را به عنوان بهترين بار ديگر پينگ‌پونك تثبيت نمايد يا آن مادر پير را خوشحال نمايد.

علي بهتر است كدام كار را انجام دهد؟

1- بازي را ببرد.

2- بازي را ببازد و آن پيرزن را خوشحال نمايد.

اين ماجرا، چه داستاني را به ياد شما مي‌آورد؟

اگر اين ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بين اين دو ماجرا را ذكر نماييد.

لطفاً موارد تفاوت این دو ماجرا را نیز ذکر نمایید.

Structure change (ΔST):

علي، معروفترین کشتيگیر زمان خود بود. علی صبح روزي که قرار بود با یک کشتيگیر جوان از استانی دیگر مسابقه بدهد در مسجد به پیرزني برخورد کرد که اینگونه براي فرزند خود دعا ميکرد «خدایا، پسرم امروز قرار است با علي مسابقه بدهد. خدایا به او کمك کن تا بتواند در این مسابقه برنده شود تا بتواند با جایزه مسابقه برای من لباسهای جدید و گران قیمت تهیه نماید.» علي با خودش فکر کرد که کشتيگیر جوان بیشتر از او به این پول نیاز دارد همچنین اگر او کشتي را از آن ورزشکار جوان ببرد قلب مادر آن جوان خواهد شکست. علی دو انتخاب داشت مسابقه را ببرد و موقعیت خود را به عنوان بهترین کشتيگیر جهان تثبیت کند یا مسابقه را ببازد و آن مادر پیر را خوشحال نماید.

علي بهتر است کدام کار را انجام دهد؟

1- بازي را ببرد.

2- بازي را ببازد و آن پیرزن را خوشحال نماید.

این ماجرا، چه داستاني را به یاد شما ميآورد؟

اگر این ماجرا، داستاني را به یاد شما ميآورد لطفاً موارد شباهت بین این دو ماجرا را ذکر نمایید.

لطفاً موارد تفاوت این دو ماجرا را نیز ذکر نمایید.

Surface + Structure change (ΔSS):

علي معروفترين بازيكن پينگ‌پونگ در شهر خود بود. صبح قبل از يك مسابقه با يك بازيكن جوان از شهر ديگر، علي در حالي‌كه در خارج از استاديوم قدم مي‌زد، به پيرزني برخورد كرد كه اينگونه براي فرزند خود دعا مي‌كرد «خدايا، پسرم امروز قرار است با علي مسابقه بدهد. خدایا به او كمك كن تا بتواند در این مسابقه برنده شود تا بتواند با جایزه مسابقه برای من لباسهای جدید و گران قیمت تهیه نماید». علي با خودش فكر كرد كه بازيكن پینگ پونگ جوان بیشتر از او به این پول نیاز دارد همچنین اگر او بازی را از آن ورزشكار جوان ببرد قلب مادر آن جوان خواهد شكست. علی دو انتخاب داشت مسابقه را ببرد و موقعیت خود را به عنوان بهترین بازیکن پینگ پونگ جهان تثبیت کند یا مسابقه را ببازد و آن مادر پیر را خوشحال نمایند.

علي بهتر است كدام كار را انجام دهد؟

1- بازي را ببرد.

2- بازي را ببازد و آن پيرزن را خوشحال نمايد.

اين ماجرا، چه داستاني را به ياد شما مي‌آورد؟

اگر این ماجرا، داستاني را به یاد شما مي‌آورد لطفاً موارد شباهت بین این دو ماجرا را ذکر نمایید.

لطفاً موارد تفاوت این دو ماجرا را نیز ذکر نمایید.

Sacred Value Change (ΔSV):

علي قرار بود با معروفترین کشتي‌گیر زمان خود کشتي بگیرد. علي صبح روز مسابقه در مسجد به پیرزني برخورد کرد که اینگونه براي فرزند خود دعا مي‌کرد «خدایا، پسرم امروز قرار است با کشتي‌گیري جوان مسابقه بدهد. خدایا به او کمك کن تا بتواند در این مسابقه برنده شود تا بتواند مقام معروفترین کشتي‌گیر زمان را تثبیت کند.» علي با خودش فکر کرد که اگر او کشتي را ببرد قلب مادر آن کشتي‌گیر خواهد شکست. علی دو انتخاب داشت مسابقه را سعي کند ببرد و خود را به عنوان بهترین کشتي‌گیر جهان تثبیت کند یا مسابقه را ببازد و آن مادر پیر را خوشحال نماید.

علي بهتر است کدام کار را انجام دهد؟

1- بازي را ببرد.

2- بازي را ببازد و آن پیرزن را خوشحال نماید.

این ماجرا، چه داستاني را به یاد شما مي‌آورد؟

اگر اين ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بين اين دو ماجرا را ذكر نماييد.

لطفاً موارد تفاوت اين دو ماجرا را نيز ذكر نماييد.

8.4.1.2.Dehghan Fadakar

Surface change (ΔSF):

مردي پس از يك روزكاري، در حالي‌كه يك چراغ قوه حمل مي‌كرد در مسير بازگشت به خانه مشاهده كرد كه به علت يك زلزله قسمتي از پل ريزش كرده است. آن مرد با اندكي دقت متوجه شد كه اتوبوسي به پل نزديك مي‌شود. او در آن شرايط دو انتخاب پيش‌رو داشت اول آنكه خود را به ايستگاه اتوبوس برساند و مدير ايستگاه را از اتفاق پيش آمده مطلع سازد و جان خود را نجات دهد و دوم آنكه با چراغ قوه خود در كنار مسير اتوبوس بايستد و با به خطر انداختن جان خود سعي نمايد راننده اتوبوس را از حادثه پيش‌رو مطلع سازد.

آن مرد بهتر است كدام كار را انجام دهد.

1- به سمت ايستگاه اتوبوس برود.

2- جان خويش را به خطر اندازد و راننده اتوبوس را مطلع نمايد.

این ماجرا، چه داستاني را به ياد شما ميآورد؟

اگر این ماجرا، داستاني را به ياد شما ميآورد لطفاً موارد شباهت بين این دو ماجرا را ذکر نماييد.

لطفاً موارد تفاوت این دو ماجرا را نيز ذکر نماييد.

Structure change (ΔST):

يك دهقان پس از يك روزكاري، در حاليكه يك چراغ نفتي حمل ميكرد، در مسير بازگشت به خانه بود که مشاهده کرد به علت ريزش سنگ از بالاي کوه، بخشي از ريل راهآهن، درست قبل از تونل از سنگ پوشيده شده است. دهقان با اندكي دقت متوجه شد که قطاري به تونل نزديك ميشود. او در آن شرايط دو انتخاب پيشرو داشت اول آنکه تلاش کند خود را به ايستگاه قطار برساند و از مدير ايستگاه بخواهد مسير قطار را عوض کند يا آنکه زندگی خود را به خطر بياندازد و در مسير قطار بايستد راننده را مطلع سازد زيرا انجام این کار باعث خواهد شد که در شهر خود معروف شده و احتمالاً یک جایزه نقدی نیز دریافت نماید.

آن مرد بهتر است کدام کار را انجام دهد.

1- به سمت ايستگاه اتوبوس برود.

2- جان خويش را به خطر اندازد و راننده اتوبوس را مطلع نمايد.

این ماجرا، چه داستاني را به ياد شما مي‌آورد؟

اگر این ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بین این دو ماجرا را ذکر نمایید.

لطفاً موارد تفاوت این دو ماجرا را نیز ذکر نمایید.

Surface + Structure change (ΔSS):

مردي پس از يك روزكاري، در حالي‌كه يك چراغ قوه حمل مي‌كرد در مسير بازگشت به خانه مشاهده كرد كه به علت يك زلزله قسمتي از پل ريزش كرده است. آن مرد با اندكي وقت متوجه شد كه اتوبوسي به پل نزديك مي‌شود. او در آن شرايط دو انتخاب پيش‌رو داشت اول آنکه خود را به ایستگاه اتوبوس برساند از مدير ایستگاه بخواهد مسير اتوبوس را عوض کند و دوم آنکه با چراغ قوه خود در كنار مسير اتوبوس بايستد. و با به خطر انداختن جان خود راننده را مطلع سازد زيرا انجام اين كار باعث خواهد شد كه در شهر خود معروف شده و احتمالاً یک جایزه نقدی نیز دریافت نماید.

آن مرد بهتر است كدام كار را انجام دهد.

1- به سمت ایستگاه اتوبوس برود.

2- جان خويش را به خطر اندازد و راننده اتوبوس را مطلع نمايد.

این ماجرا، چه داستاني را به یاد شما مي‌آورد؟

اگر این ماجرا، داستاني را به یاد شما مي‌آورد لطفاً موارد شباهت بین این دو ماجرا را ذکر نمایید.

لطفاً موارد تفاوت این دو ماجرا را نیز ذکر نمایید.

8.4.1.3. Hossein Fahmide

Surface change (ΔSF):

در طي جنگ بوسنیایي‌ها و صربها یك پسر جوان كه به سختي خود را وارد نیروهاي مسلح كرده بود، یك روز در هنگام جنگ با یك كاروان از اتوبوسهاي دشمن مواجه شد كه سرباز و سلاح حمل مي‌كردند. اگر این اتوبوسها متوقف نمي‌شدند به دشمن كمك مي‌كردند كه بخشي از شهریرا كه آن پسر جوان در آن مي‌جنگید ویران نمایند. آن پسر جوان در آن شرایط دو انتخاب پیش‌رو داشت اول آنكه فرمانده خود را مطلع سازد تا از واحدهاي دیگر كمك گرفته شود تا آن كاروان متوقف گردد و دوم آنكه خود را با سرعت به زیر اتوبوس اول بیاندازد و با به كار انداختن یك مین از كار افتاده و كشتن خویش، با انهدام این اتوبوس مانع ادامه حركت كاروان گردد.

در این شرایط بهتر است آن پسر جوان كدام كار را انجام دهد؟

1- خود را به فرمانده رسانده و او را مطلع کند.

2- زندگي خويش را براي متوقف كردن آن كاروان قرباني نمايد.

اين ماجرا، چه داستاني را به ياد شما مي‌آورد؟

اگر اين ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بين اين دو ماجرا را ذكر نماييد.

لطفاً موارد تفاوت اين دو ماجرا را نيز ذكر نماييد.

Structure change (ΔST):

در طي یک جنگ پسر جوانی که خود را به سختي خود را در ميان نيروهاي مسلح وارد كرده بود، با يك كاروان از تانك ها مواجه شد كه در صورتيكه از حركت نمي‌ايستاد قسمتي از شهري را كه آن پسر در آن مي‌جنگيد ويران مي‌كردند. آن پسر مي‌توانست تلاش كند تا خود را به فرمانده برساند و او را در مورد شرايط مطلع سازد تا از واحدهای ديگر کمک گرفته شود تا آن كاروان متوقف شود و يا مي‌توانست با قرباني كردن خويش جلوي حركت تانكها را بگيرد.

در این شرایط بهتر است آن پسر جوان کدام کار را انجام دهد؟

1- خود را به فرمانده رسانده و او را مطلع کند.

2- زندگي خويش را براي متوقف كردن آن كاروان قرباني نمايد.

این ماجرا، چه داستاني را به ياد شما مي‌آورد؟

اگر این ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بين اين دو ماجرا را ذكر نماييد.

ـ لطفاً موارد تفاوت اين دو ماجرا را نيز ذكر نماييد.

Surface + Structure change (ΔSS):

در طي جنگ بوسنيايي‌ها و صربها يك پسر جوان كه به سختي خود را وارد نيروهاي مسلح كرده بود، يك روز در هنگام جنگ با يك كاروان از اتوبوسهاي دشمن مواجه شد كه سرباز و سلاح حمل مي‌كردند. اگر اين اتوبوسها متوقف نمي‌شدند آنها به دشمن كمك مي‌كردند كه بخشي از شهري كه آن پسر جوان در آن مي‌جنگيد را ويران نمايند. آن پسر جوان در آن شرايط دو انتخاب پيش‌رو داشت اول آنكه فرمانده خود را مطلع سازد تا از واحدهاي ديگر كمك گرفته شود تا آن كاروان متوقف گردد و دوم آنكه خود را با سرعت به زير اتوبوس اول بياندازد و با به كار انداختن يك مين از كار افتاده و كشتن خويش، با انهدام اين اتوبوس مانع ادامه حركت كاروان گردد.

در اين شرايط بهتر است آن پسر جوان كدام كار را انجام دهد؟

1- خود را به فرمانده رسانده و او را مطلع كند.

2- زندگي خويش را براي متوقف كردن آن كاروان قرباني نمايد.

اين ماجرا، چه داستاني را به ياد شما مي‌آورد؟

ـ اگر اين ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بين اين دو ماجرا را ذكر نماييد.

ـ لطفاً موارد تفاوت اين دو ماجرا را نيز ذكر نماييد.

Sacred Value Change (ΔSV):

در طي جنگ بوسنيايي‌ها و صربها يك پسر جوان صربي كه به سختي خود را وارد نيروهاي مسلح كرده بود، يك روز در هنگام جنگ با يك كاروان از اتوبوسهاي بوسنيايي‌ها مواجه شد كه سرباز و سلاح حمل مي‌كردند. اگر اين اتوبوسها متوقف نمي‌شدند به بوسنيايي‌ها كمك مي‌كردند كه بخشي از شهري را كه آن پسر جوان در آن مي‌جنگيد ويران نمايند. آن پسر جوان در آن شرايط دو انتخاب پيش‌رو داشت اول آنكه فرمانده خود را مطلع سازد تا از واحدهاي ديگر كمك گرفته شود تا آن

کاروان متوقف گردد و دوم آنکه خود را با سرعت به زیر اتوبوس اول بیاندازد و با به کار انداختن يك مين از كار افتاده و کشتن خویش، با انهدام این اتوبوس مانع ادامه حرکت کاروان گردد.

در این شرایط بهتر است آن پسر جوان کدام کار را انجام دهد؟

1- خود را به فرمانده رسانده و او را مطلع کند.

2- زندگي خويش را براي متوقف كردن آن كاروان قرباني نمايد.

اين ماجرا، چه داستاني را به ياد شما مي‌آورد؟

اگر اين ماجرا، داستاني را به ياد شما مي‌آورد لطفاً موارد شباهت بين اين دو ماجرا را ذكر نماييد.

لطفاً موارد تفاوت اين دو ماجرا را نيز ذكر نماييد.

8.4.2. English version of the stimuli for the control group

8.4.2.1. Pourya Vali

Surface change (ΔSF):

John is the most famous swimmer of his city. The morning before a match with a young athlete from another city, he goes for a walk outside the stadium and sees the mother of the young athlete praying and saying "God, my son is going to play a match with John the famous swimmer. Please watch over him and help him win the match so he can use the prize money to get married". John has two choices, he can either win the match and keep his status as the best swimmer or he could lose the match and make the old mother happy.

What should John do?

a. Win the match

b. Lose the match and make the old woman happy

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Structure change (ΔST):

Matt is the most famous tennis player of his time. The morning before a match with a young athlete from another country, he goes for a walk outside the stadium and sees the mother of the young athlete praying and saying “God, my son is going to have a match with Matt the famous tennis player. Please watch over him and help him win the match so he can use the prize money to buy me new expensive cloths”. Matt has two choices, he can either win the match and keep his status as the number one tennis player in the world or he could lose the match make the old mother happy.

What should Matt do?

a. Win the match

b. Lose the match and make the old woman happy

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Surface + Structure change (ΔSS):

John is the most famous swimmer of his city. The morning before a match with a young athlete from another city, he goes for a walk outside the stadium and sees the mother of the young athlete praying and saying "God, my son is going to play a match with John the famous swimmer. Please watch over him and help him win the match so he can use the prize money to buy me new expensive cloths". John has two choices, he can either win the match and keep his status as the best swimmer or he could lose the match and make the old mother happy.

What should John do?

a. Win the match

b. Lose the match and make the old woman happy

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Sacred Value change (ΔSV):

Andrew was going to play a match against the most famous tennis player of his time. The morning before the match, Andrew goes for a walk outside the stadium and sees the mother of the famous athlete praying and saying “God, my son is going to have a match with the young athlete Andrew. Please watch over him and help him win the match so he can keep his status as the best player”. Andrew has two choices, he can either win the match and beat the number one player in the world or he could lose the match make the old mother happy.

What should Andrew do?

a. Win the match

b. Lose the match and make the old woman happy

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

8.4.2.2. Dehghan Fadakar

Surface change (ΔSF):

A man is going to work carrying a flashlight. He notices that as the result of an earthquake, a bridge has collapsed. He walks passed the bridge and realizes that a bus is heading towards the tunnel. He has two options: he can either try to run to the station on time, inform the station manager and save his own life, or he can use his flashlight, stand in the way of the of the bus, risk his life and try to signal the bus.

What should the man do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Structure change (ΔST):

A farmer is returning home from a day of work carrying an oil lamp. He notices that as the result of a landslide, parts of a railroad just outside of a tunnel has been covered with stones. He walks passed the tunnel and realizes that a train is heading towards the tunnel. The farmer has two options, he can either try to run to the station on time and have the station manager reroute the train, or risk his life, by standing on the tracks, which will make him famous in his town and he would potentially receive a cash prize.

What should the man do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Surface + Structure change (ΔSS):

A man is going to work carrying a flashlight. He notices that as the result of an earthquake, a bridge has collapsed. He walks passed the bridge and realizes that a bus is heading towards the tunnel. He has two options: he can either try to run to the station on time and have the station manager reroute the train, or he can use his flashlight, stand in the way of the of the bus, risk his life and try to signal the bus, which will make him famous in his town and he would potentially receive a cash prize.

What should the man do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

8.4.2.3. Hossein Fahmide

Surface change (ΔSF):

During the Bosnia and Serbian war, a young boy sneaks in to the army. One day during the war, he is confronted with a convoy of enemy buses carrying soldiers and weapons. If these buses are not stopped, they will help the enemy destroy part of the city that the boy is fighting at. He can either try to run to his commander on time, inform him about the situation and save his own life or he can stop a bus by running underneath it and activating a mine which otherwise would not work.

What should the young boy do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Structure change (ΔST):

During a war, a young boy who has sneaked into the army, is confronted with a tank that if not stopped will destroy a part of the city that the boy is fighting at. He can either try to run to his commander on time and inform him about the attack which would cause the commander to issue a strike from other units against the tanks or he can stop one tank by sacrificing his own life.

What should the young boy do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Surface + structure change (ΔSS):

During the Bosnia and Serbian war, a young boy sneaks in to the army. One day during the war, he is confronted with an enemy bus carrying soldiers and weapons. If this bus is not stopped, it will help the enemy destroy part of the city that the boy is fighting at. He can either run to his commander on time, inform him about the situation which would cause the commander to issue a strike from other units against the convoy of buses or he can stop a bus by running underneath it and activating a mine which otherwise would not work.

What should the young boy do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

Sacred Value change (ΔSV):

During the war between Muslim Bosnians and Serbians, a young Serbian boy sneaks in to the army. One day during the war, he is confronted with a convoy of Bosnian buses carrying soldiers and weapons. If these buses are not stopped, they will help the enemy destroy part of the city that the boy is fighting at. He can either try to run to his commander on time, inform him about the situation and save his own life or he can stop a bus by running underneath it and activating a mine which otherwise would not work.

What should the young boy do?

a. Run away

b. Sacrifice his own life

What narrative does this scenario remind you of?

If it reminds you of any narratives, please list the similarities between the two.

Please list the differences between the two.

8.5. Appendix E: An interview script

Open-ended Task Script:

a) Imagine that the population of the bears/poplar trees in the local forest disappeared/doubled. We don't know why or how this happened, we just know they have disappeared/doubled. Now imagine that it is one year later since all of the bears/poplar trees disappeared/doubled. Do you think that the disappearance/multiplying of the bears would have an effect on other plants in the forest? On any animals?

[If they say 'yes', move on to (b)]

[If they say 'no', move on to (c)]

b) In what way? Can you provide any specific examples of plants and animals that would be affected?

[If they cannot provide examples, probe for the following]:

- squirrels
- coyotes
- deer
- eagles
- berries
- trees

c) What do you think the situation would look like 30 years after the disappearance/multiplying of the bears? Do you think the situation would look different 30 years later compared to 1 year later?

[If they say yes, ask ‘How?’]