MoralDM: A Computational Model of Moral Decision-Making

Article · April 2012

CITATIONS
4

READS
636

5 authors, including:

Morteza Dehghani
University of Southern California
134 PUBLICATIONS **2,900** CITATIONS
SEE PROFILE

Emmett Tomai
University of Texas - Pan American
27 PUBLICATIONS **370** CITATIONS
SEE PROFILE

Ken Forbus
Northwestern University
65 PUBLICATIONS **1,124** CITATIONS
SEE PROFILE

Rumen Iliev
University of Michigan
24 PUBLICATIONS **1,538** CITATIONS
SEE PROFILE

MoralDM: A Computational Model of Moral Decision-Making

Morteza Dehghani[1], Emmett Tomai[1], Ken Forbus[1], Rumen Iliev[2], Matthew Klenk[1]
{morteza, etomai, forbus, r-iliev, m-klenk}@northwestern.edu

[1]Qualitative Reasoning Group, Northwestern University
2145 Sheridan Road, Evanston, IL 60208-0834, USA

[2]Department of Psychology, Northwestern University
2029 Sheridan Road, Evanston, IL 60208-2710, USA

Abstract

We present a cognitively motivated model of moral decision-making, MoralDM, which models psychological findings about utilitarian and deontological modes of reasoning. Current theories of moral decision-making extend beyond pure utilitarian models by including contextual factors that vary culturally. Our model employs both first-principles reasoning and analogical reasoning to implement rules of moral decision-making and compare previously solved cases to novel situations. The different impacts of secular versus sacred values are modeled via qualitative reasoning, using an order of magnitude representation. We evaluate MoralDM on stimuli taken from two psychology experiments.

Keywords: Decision Making; Analogical Reasoning, Order of Magnitude Reasoning, Natural Language Processing

Introduction

Traditionally, models of decision-making only concentrated on the utility of outcomes calculated using axioms of economic theory. Recent psychological results have shed light on the process of human decision-making by showing predictable violations of these axioms (Kahneman, Slovic, and Tversky, 1982). One of the domains in which traditional normative utilitarian models fail to predict human behavior is the domain of moral reasoning.

Psychological evidence indicates that people facing moral dilemmas often do not act in utilitarian ways. Baron and Spranca (1997) suggested the existence of *protected values*, which are not allowed to be traded-off, regardless of the consequences. Further, they suggest that these protected values "arise out of deontological rules about actions rather than outcomes". A similar trade-off blockage was proposed by Tetlock (2000), who defined *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". Consider the traffic scenario (from Ritov & Baron 1999) below:

> A program to combat accidents saves 50 lives per year in a specific area. The same funds could be used to save 200 lives in another area, but the 50 lives in the first area would be lost.
>
> Do you transfer the funds?

While the utilitarian decision would transfer funds to the second area, the majority of the participants choose to not transfer them. People who have sacred values tend to reject trade-offs and often show strong emotional reactions, such as anger, when these values are challenged.

This paper proposes a cognitive model of moral decision-making, called MoralDM, which models psychological findings about utilitarian and deontological modes of reasoning. MoralDM uses both first-principles and analogical reasoning to implement rules of moral decision-making and utilize previously made decisions. The impacts of secular versus sacred values are modeled via qualitative reasoning, using an order of magnitude representation. We test this model on stimuli from two psychology experiments. To reduce tailorability, we use a natural language understanding system to assist in producing formal representations from the stimuli re-rendered in simplified English.

We first discuss the role of analogy in the process of human decision-making. Next, we give a brief overview of protected values and quantity insensitivity and how qualitative reasoning can be useful in calculating utilities in a cognitively plausible way. Then, we discuss how MoralDM works. Finally, we describe experimental results and discuss future work.

Decision-Making and Analogy

The link between analogy and decision-making has been explored from various perspectives, including consumer behavior (Gregan-Paxton, 1998), political reasoning (May, 1973) and legal decision-making (Holyoak and Simon, 1999). 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 and Medin, 2002). 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 go into 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 and Taber, 1990; Markman and Moreau, 2001).

To model analogy in decision-making, we use the Structure-Mapping Engine (SME) (Falkenhainer et al. 1989), a computational model of similarity and analogy based on Gentner's (1983) structure mapping theory of analogy in humans. 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 the maximal structurally consistent mapping between the cases. This mapping consists of a set of *correspondences* between entities and expressions. 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 target. Mappings also include *candidate inferences*, conjectures about the target using expressions from the base that, while unmapped in their entirety, have subcomponents that participate in the mapping's correspondences.

Sacred Values and Quantity Insensitivity

In the presence of sacred values, people tend to be less sensitive to outcome utilities in their decision-making. This results in decisions which are contrary to utilitarian models. We claim that this can be modeled by using existing qualitative reasoning formalisms. After summarizing the relevant moral decision-making findings, we present a simplified implementation of Dague's (1993) ROM(R) qualitative order of magnitude formalism which we use to capture these results.

Sacred or protected values concern acts and not outcomes. When dealing with a case involving a protected value, people tend to be concerned with the nature of their action rather than the utility of the outcome. 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. The amount of sensitivity (or insensitivity) towards outcomes vary with the context. For example, Bartels and Medin (2007) argue that the agent's sensitivity towards the outcome of a moral situation is dependent on the agent's focus of attention. Lim and Baron (1997) show that people's sensitivity towards outcomes varies across cultures. They show that people in different cultures tend to protect different values and also the degree of sensitivity towards a certain protected value is different 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 act less quantity sensitive when their action directly, rather than indirectly, causes harm.

We model quantity sensitivity by using Dague's (1993) ROM(R) qualitative order of magnitude formalism. Order of magnitude representations provide the kind of stratification that seems necessary for modeling the impact of sacred values on reasoning. One of the features of ROM(R) is that it includes a parameter, k, which can be varied to capture differences in quantity sensitivity. We implemented a simplified version of ROM(R) using one degree of freedom, k, resulting in three binary relations which 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, with a higher k resulting in less quantity sensitivity. Depending on the sacred values involved and the causal structure of the scenario, we vary k to capture sensitivity towards the utility of the outcome.

MoralDM

Our model of moral decision-making, MoralDM, has been implemented using the FIRE reasoning engine and its underlying knowledge base. The knowledge base contents are a 1.2 million fact subset of Cycorp's ResearchCyc knowledge base, which provides formal representations about everyday objects, people, events and relationships. The KB also includes representations we have developed to support qualitative and analogical reasoning. The scenarios, decisions and rule sets used in MoralDM are all represented uniformly and stored in this KB.

MoralDM operates in two mutually exclusive modes of decision-making: utilitarian and deontological. If there are no sacred values involved in the case being analyzed, MoralDM applies traditional rules of utilitarian decision-making by choosing the action which provides the highest outcome utility. On the other hand, if MoralDM determines that there are sacred values involved, it operates in deontological mode and becomes less sensitive to the outcome utility of actions, preferring inactions to actions.

Moral decision-making dilemmas are represented in predicate calculus. Figure 1 contains the representation for the choice between ordering the transfer of funds and inaction from the traffic scenario presented at the beginning of this paper. The Order of Magnitude Reasoning (OMR) module calculates the relationship between the utility of each choice. Using the outcome of OMR, MoralDM utilizes a First-Principles Reasoning (FPR) module and an Analogical Reasoning (AR) module to arrive at a decision. FPR suggests decisions based on rules of moral reasoning.

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

Figure 1: Predicate calculus used to represent the decision in the traffic scenario between inaction and ordering the transfer of funds

AR compares a given scenario with previously solved decision cases to suggest a course of action. We believe using both techniques gives the system more power to explain different decision-making scenarios and provides a more cognitively plausible approach to decision-making. Our 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. Figure 2 depicts the MoralDM Architecture.

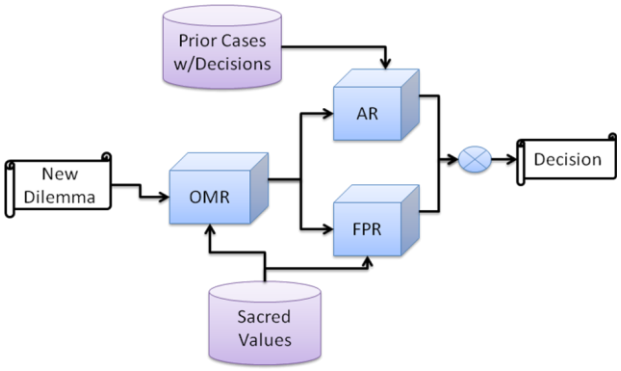

Figure 2: MoralDM Reasoning Architecture

FRP and AR 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 results of the modules do not agree, MoralDM reports the results of the FPR module. If both fail, the system is incapable of making a decision. Next, we discuss each module in detail.

Order of Magnitude Reasoning Module

OMR uses the sacred values for the culture being modeled and the causal structure of the scenario to determine the order of magnitude relationship of the outcome utilities. Using the predicate calculus description of the scenario, OMR calculates the expected utility of each choice by summing the utility of its consequences. For each consequence of a choice, OMR ascertains if the outcome is positive or negative and identifies any sets whose cardinality matters in the decision (e.g., number of people at risk).

After computing utilities, OMR selects a *k* value. For detailed analysis of how *k* is computed please refer to Dehghani et al. (2008). If the decision involves a sacred value for the modeled culture, a higher *K* value will be used. This can change the relationship between utilities from greater than to almost equal, resulting in the system being less sensitive to the numeric utility of the outcome. On the other hand, if there are no sacred values involved, the system substitutes lower values for *k* thereby making the system more quantity sensitive to the computed utilities. In addition to sacred values, the causal structure of the scenario affects *k*. OMR checks to see if the scenario contains patient intervention or agent intervention. It uses low quantity sensitivity for agent intervention and high otherwise, consistent with psychological findings (Waldmann and Dieterich 2007). The system also checks for direct versus indirect causation. In the case of indirect causation, a lower degree of sensitivity is applied.

Returning to the traffic scenario, there are two choices: transferring funds and inaction. For transferring funds, there are two consequences: 200 lives in the second area will be saved while 50 people in the first area will die. Consulting the KB, the system determines that dying has negative utility and saving positive, resulting in a choice utility of 150 for the transferring choice. Using the same procedure, the utility for inaction is calculated to be -150. Given that both choices involve agent intervention and indirect causation, there are no structural differences between the two choices. Therefore, the *k* value is determined solely by the existence of sacred values. In this case, causing someone to die is a sacred value. Using ROM(R), the relationship between the utilities of the two choices is calculated to be almost equal. On the other hand, if there had not been a sacred value, a smaller *k* would have been chosen causing the relationship between the utilities to be an order of magnitude greater. These utilities (150 and -150) and the computed relationship (almost equal) are provided to FPR and AR.

First-Principles Reasoning Module

Motivated by moral decision-making research, FPR makes decisions based upon sacred values, computed utilities, action vs. inaction and the orders of magnitude relationship between utilities. FPR uses two modes for making decisions, utilitarian and deontological. The utilitarian mode, which 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 sacred values and without an order of magnitude difference between outcomes, the deontological mode is invoked and the choice that does not violate a sacred value is selected.

In the traffic scenario, there is a sacred value, people dying, and no order magnitude difference between the

utility of the two choices. Therefore, FPR uses the deontological mode to select the inaction choice.

These methods are mutually exclusive, returning at most one choice per scenario. Given the breadth of moral reasoning scenarios, the rules implementing FPR are not complete. Therefore, FPR necessarily fails on some scenarios. These cases highlight the need for the hybrid-reasoning approach taken in MoralDM.

Analogical Reasoning Module

Running concurrently with FPR, AR uses comparisons between new cases and previously solved cases to suggest decisions. When faced with a moral 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. 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. If this score is higher than a certain threshold and both scenarios contain the same order of magnitude relationship between outcome utilities, then the candidate inference indicating which choice to select is considered a valid analogical decision. If the scenarios have different order of magnitude relationships, it is likely that a different mode of reasoning should be used for the target scenario. In this case, AR rejects the candidate inference. After comparing against all of the solved scenarios, AR selects the choice with the highest number of 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 traffic scenario example, AR can solve this decision problem through an analogy with a starvation scenario given below, in which the system chose to not order the convoy to go to the second camp:

> 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?

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 starvation scenario contains the same the order of magnitude relationship (almost equal) as the transfer of funds scenario, the system accepts the analogical decision.

Evaluation

We evaluated MoralDM using moral decision-making scenarios from two psychology studies. We used the Explanation Agent Natural Language Understanding system (EA NLU, Kuehne & Forbus, 2004) to produce predicate calculus descriptions from simplified English versions of the stimuli. Unrestricted automatic natural language understanding is currently beyond the state of the art. Consequently, EA NLU uses a controlled language and operates semi-automatically, enabling experimenters to select among options presented by the system. Our use of controlled language is inspired by both CMU's KANT project (cf. Mitamura & Nyberg 1995) and Boeing's controlled language work (cf. Clark et al. 2005). This practical approach allows us to broadly handle syntactic and semantic ambiguities and to build deep representations suitable for complex reasoning.

The first experiment includes the 4 decision-making scenarios, each describing two outcomes, from Waldmann and Dieterich's (2007) experiments. The second experiment contains the 12 scenarios from Ritov and Baron's (1999) paper. In all these decision-making scenarios, traditional utility theories fail to predict the subjects' responses. We compare MoralDM's decisions to subjects' responses in these experiments as reported by the authors. When MoralDM's decision matches those of the majority of subjects, we consider it a correct choice.

The AR module requires previously solved decision cases to act as past experience to draw from. In each experiment there are *n* total scenarios. For each scenario, a library of solved decision cases consisting of the *n-1* other scenarios was made available to the system. Therefore, in the first experiment, the AR module always compared the decision scenario against 3 solved cases and in the second experiment, it always compared the scenario to 11 solved cases.

Experiment 1

In the first experiment, we tested our system on all of the scenarios examined by Waldmann and Dieterich (2007). For each scenario, subjects were asked to choose between two outcomes which offer identical outcome utilities but have different causal structures. More specifically, in one of the cases the focal action is performed on the agent of harm, whereas in the other case, the action influences a potential patient. Here is one of the scenarios from Waldmann and Dieterich (2007) study, where subjects were asked to choose between the two cases:

> 1. In a restaurant, a bomb threatens to kill 9 guests. The bomb could be thrown onto the patio, where 1 guest would be killed.
>
> 2. In a restaurant, a bomb threatens to kill 9 guests. One guest could be thrown on the bomb, which would kill this 1 guest.

The first case is the agent-intervention variant case and the second patient-intervention. The authors report that subjects were more likely to choose the first variant over the second one.

To reduce tailorability, all the scenarios in this experiment were translated into predicate calculus using EA NLU.

	# of correct decisions
MoralDM	4 (100%)
First-principles	4 (100%)
Analogical Reasoning	3 (75%)

Table 1: MoralDM results

MoralDM's decisions followed the subjects' answers in all four of the scenarios. Table 1 contains the results for MoralDM and each of the reasoning modules. In 3 scenarios, both first-principles reasoning and analogical reasoning provide the correct answer. Analogical reasoning selected the wrong answer in one scenario. This particular scenario had a different causal structural from the other cases. First-principles reasoning answered correctly in all of the scenarios.

Experiment 2

In this experiment, all 12 moral decision-making scenarios from Ritov and Baron (1999) were used as inputs. These scenarios cover a wide range of topics such as civil rights, nature preserves, combating traffic accidents, Jewish settlements, Arab villages, etc. After reading each scenario, subjects were asked to choose between two choices. The outcome utilities of the two choices are the same, but the type of actions and events involved in reaching the outcomes are different.

We used a combination of EA NLU and manual encoding to translate these scenarios into predicate calculus during ongoing development of the EA NLU system.

	# of correct decisions
MoralDM	11 (92%)
First-principles	8 (67%)
Analogical Reasoning	11 (92%)

Table 2: MoralDM results

Out of the 12 scenarios, MoralDM makes decisions matching those of participants on 11 scenarios. Table 2 displays the results of MoralDM and the FPR and AR modules. In 8 scenarios, both first-principles reasoning and analogical reasoning provide the correct answer. In three scenarios, first-principles reasoning fails to make a prediction, but analogical reasoning provides the correct answer. In 1 scenario, both reasoning strategies fail.

Analogical reasoning fails on one of the cases because the causal structure of the case is very different from the other cases. First-principles reasoning fails in four of the scenarios because these cases require special rules for dealing with their unique structure or content. We believe writing special rules for cases is not cognitively plausible and we currently working on methods for deriving these rules automatically using analogical reasoning.

Discussion

We evaluated MoralDM on all the stimuli from two moral decision-making experiments. Of the 16 different moral scenarios, MoralDM's decisions matched the subjects' responses on 15 scenarios. These results provide evidence for MoralDM as a model for moral decision-making. They support the claim that an order of magnitude representation is effective for modeling people's sensitivity to outcome quantities in decision scenarios. Moreover, the hybrid-reasoning approach allowed MoralDM to solve more scenarios than either first-principles reasoning or analogical reasoning alone. Frequently, when one module failed to arrive at a decision, the other module arrived at the correct decision. Part of our future work is to further explore the interaction between these two types of reasoning.

Conclusions and Future Work

We presented MoralDM, a computational model of moral decision-making which captures psychological results concerning deontological and utilitarian modes of reasoning. MoralDM uses a qualitative order of magnitude representation to model quantity sensitivity concerning outcome utilities, and a combination of first-principles and analogical reasoning to make decisions. To reduce tailorability, we used EA NLU to produce formal representations of stimuli from natural language.

Computational models of cultural reasoning are receiving increasing attention. For example, 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". CARA uses semantic web technologies and opinion extraction from weblogs to build cultural decision models consisting of qualitative rules and utility evaluation. While we agree that qualitative reasoning must be integrated with traditional utility evaluation, we also believe that analogy plays a key role in moral reasoning. Moreover, we differ by evaluating our system against psychological studies.

We plan to test MoralDM on a wider range of problems and use it to model decision-making in a variety of different cultures. 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. By focusing on decisions from a specific culture, we can explore automatic model construction for making novel predictions about the behavior of a certain group (Dehghani et al. 2007). As our decision libraries for various cultural groups grow, we will incorporate MAC/FAC (Forbus et al., 1995) as a cognitively plausible model of retrieving relevant precedents.

Acknowledgments

This research was supported by a grant from an AFSOR MURI. We thank Andrew Lovett for many useful insights.

References

Baron, J., and Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes* 70, pp. 1–16.

Bartels, D. M. and Medin, D. L. (2007) Are Morally-Motivated Decision Makers Insensitive to the Consequences of their Choices? *Psychological Science*, 18, 24-28.

Clark, P., Harrison, P. and Thompson, J. (2003). A Knowledge-Driven Approach to Text Meaning Processing. *Proceedings of the HLT Workshop on Text Meaning Processing*, pp 1-6.

Dauge, P. (1993). Symbolic Reasoning with Relative Orders of Magnitude. In *Proceedings of the 13th IJCAI*.

Dehghani, M., Tomai E., Forbus, K., Klenk, M. (2008). Order of Magnitude Reasoning in Modeling Moral Decision-Making. To appear in *22nd International Workshop on Qualitative Reasoning*.

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

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

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

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

Glad, B., and Taber, C. S. (1990). Images, learning and the decision to use force: The domino theory of the United States. In Glad, B., editors, *Psychological dimensions of war*, pp.56-82. Sage Publications, Newbury Park, CA.

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.

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

Kahneman, D., Slovic, P., Tversky, A., Eds. Judgment Under Uncertainty: Heuristics and Biases (Cambridge Univ. Press, New York, 1982)

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

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

Markman, A and Medin, D.L. (2002). *Decision Making*. Stevens Handbook of Experimental Psychology, 3rd edition: Volume 2, Memory and Cognitive Processes. New York: Wiley.

Markman, A.B., and Moreau, C. P. (2001). Analogy and analogical comparison in choice. In Gentner, D., Holyoak, K. J., and Kokinov, K. J., editors, *Analogy: Theoretical and Empirical Research*. pages 363-400. The MIT Press, Cambridge, MA.

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

Mitamura, T., and Nyberg, E. H. (1995). Controlled English for Knowledge-Based MT: Experience with the KANT System. In *Proceedings of 6th International Conference on Theoretical and Methodological Issues in Machine Translation*, Leuven, Belgium.

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

Subrahmanian, VS., Albanese, M., Martinez, V., Nau, D., Reforgiato, D., Simari, G., Sliva, A., and WIlkenfeld, J. 2007. CARA: A Cultural Adversarial Reasoning Architecture. *IEEE Intelligent Systems.* 22(2): 12-16.

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

Waldmann, M. R., and 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.

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