Title

Using Qualitative Reasoning for the Attribution of Moral Responsibility

Permalink
https://escholarship.org/uc/item/4d6950x2

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN
1069-7977

Authors
Tomai, Emmett
Forbus, Ken

Publication Date
2008

Peer reviewed

Using Qualitative Reasoning for the Attribution of Moral Responsibility

Emmett Tomai **Ken Forbus**

Qualitative Reasoning Group, Northwestern University, 2133 Sheridan Rd, Evanston, IL 60208
{etomai,forbus}@northwestern.edu

Abstract

We present a computational model, based on Attribution theory, of responsibility judgment for negative events. Our model uses Qualitative Process theory to reason over the continuous parameters involved in attribution, avoiding the need for ad-hoc assignment of quantitative values. Qualitative reasoning allows our model to infer relative amounts of responsibility for a situation in a manner that is consistent with relative amounts of blame attributed in a psychological experiment by Mao and Gratch [Mao & Gratch 2005].

Who is to blame?

Bad things happen, and blame quickly follows. From the affairs of nations to personal misfortunes, accountability is an important part of how we understand the world around us. But how does one go from perceiving situations to judging responsibility? This question has been the topic of much research in social psychology. Recently, efforts have been made to create computational models that capture the process of responsibility judgment.

This paper describes how Qualitative Process theory [Forbus 1984] can be used in such modeling. We briefly summarize aspects of Attribution theory relevant to responsibility and blame judgments, then discuss the Mao and Gratch computational model [Mao & Gratch 2005][Mao 2006]. We present an alternative model for attribution of blame based on QP theory, which we claim better represents the underlying theory. Experimental results using data collected by Mao show that our model captures that data better, and makes additional predictions.

Attribution Theory

The goal of Attribution theory [Heider 1958] is to identify the conditions that will lead a perceiver, through an *attribution process*, to attribute some behavior, event or outcome to an internal disposition of the agent involved, as opposed to an environmental condition. Attributions depend on the perceiver's knowledge. Attribution of blame has been addressed by Shaver [1985] and Weiner [1995]. Shaver distinguishes *cause*, *responsibility* and *blameworthiness*. For a given negative outcome, cause is defined as being an insufficient but necessary part of a condition, which is itself unnecessary but sufficient for that result. The theory only covers causes which represent human agency. Responsibility is "moral accountability", distinct from legal responsibility or the responsibilities of a formal office. Blame is moral condemnation that follows from responsibility for a morally reprehensible outcome.

Shaver's attribution process begins with a negative outcome and assigns responsibility to an involved agent by sequentially evaluating five dimensions: *causality, intentionality, coercion, appreciation,* and *foreknowledge*.

Causal involvement in the negative outcome is a prerequisite for any responsibility to be assigned. Shaver characterizes intention as a scale of deliberateness with intentional at one end and involuntary at the other, such that the highest degree of intention should result in the strongest judgment of responsibility. Intention, however, can be moderated by coercion and appreciation. Coercion captures the force exerted by another agent which limits the available choices, from a social standpoint, for the agent in question. This could be through some direct threat or via an authority relationship. An agent who is coerced is assigned less responsibility than one who acts intentionally in the absence of coercion. Appreciation concerns the perceiver's judgment as to whether the agent in question has the capacity to understand that the outcome in question is morally wrong. If the agent does not have such capacity, they still bear some responsibility but are held exempt from blame. Foreknowledge is defined as the extent to which the agent was aware that an action would result in the outcome, prior to execution. Again, it is the perceiver's judgment of the knowledge the agent possessed that is evaluated. In the absence of intentionality, Shaver attributes responsibility based on foreknowledge.

In Shaver's model foreknowledge may be what the agent is thought to know (epistemic) or what the perceiver thinks the agent should have known (expected). However, it says little about the contribution of expected foreknowledge. This is not surprising as his model focuses on the perception of the agent's deliberative process. Weiner's model [Weiner 1995], by contrast, focuses on attribution of responsibility in cases of achievement and failure. In the case where an agent has failed to have expected foreknowledge, this model predicts that the perception of *causal controllability* over that failure determines the degree of responsibility attributed.

Blame in Shaver's model follows from responsibility unless there is a justification or excuse. Justification does not deny responsibility; instead it is an argument about why blame should not be assigned despite responsibility. An example would be when someone shot someone else dead, but did it in self-defense. Excuses deny responsibility by appealing the judgments of the dimensions (e.g. "I didn't know", "I didn't mean it"). Successful intervention by an excuse alters the assignment of responsibility.

Mao and Gratch Computational Model

Mao, in collaboration with Gratch [Mao 2006][Mao & Gratch, 2005] developed a computational model of responsibility assignment which models the judgments of *attribution variables* based on the dimensions of causality, intentionality, coercion and foreknowledge, and the attribution of blame[1] following from those judgments. It does not deal with justifications and excuses, thus blame follows directly from responsibility.

Mao's work is an important step towards modeling blame attribution. However, there are three limitations we address here. First, as [Mao 2006] observes, it uses Boolean values for attribution variables, whereas Attribution theory describes the dimensions of responsibility in terms of scalar values. Second, all blame is assigned to a single agent (or group of agents in a joint action). This is inconsistent with the human data in Mao's own experiment. Third, the degree of blame assigned by the system is limited to a value of *high* for intentional action and a value of *low* in the absence of intention. These assignments also do not match up with her data.

Qualitative Model of Attribution

We claim that these limitations can be addressed by encoding Attribution theory in Qualitative Process (QP) theory [Forbus 1984]. We claim that this model makes more informative distinctions between blame assignments both within and across scenarios.

While physical domains have been a major focus of QR research, researchers are increasingly finding QR techniques useful in fields where theories are expressed in continuous parameters more generally, including organization theory (cf. [Kamps & Peli, 1995]), economics (cf. [Steinmann, 1997]), and political reasoning (cf. [Forbus & Kuehne 2005]). Qualitative reasoning, we believe, provides an especially appropriate level of representation for reasoning about social causality. Theories typically are expressed in terms of continuous parameters, such as "amount of intention" and "degree of foreknowledge", but there tend to not be principled ways to move to quantitative models and numerical values for such parameters. In those circumstances, qualitative modeling is a more rigorous way to proceed, and ordinal fitting with human data becomes the most robust measure.

Attributing dimensions of responsibility

We represent attribution variables for intentionality, coercion and foreknowledge as nonnegative continuous parameters. Judgments of causality remain Boolean, as that is the extent of their impact in Shaver's model. The dimension of appreciation is not addressed by this model. A value of zero is a lower limit point indicating the absence of responsibility, intentionality, coercion or foreknowledge in the judgment of the perceiver. Given a scenario involving a negative outcome and some number of agents involved, our model attributes qualitative values and constraints to these variables according to evidence from observable events in the scenario.

A significant distinction is made between evidence of act and outcome intention, following [Weiner 2001]. It is assumed that an agent intends any action that they perform or orders performed. If the action is known by the agent to have only one outcome, then that outcome is also intended. There is considerable philosophical discussion on whether foreknowledge of multiple outcomes implies intention of all those outcomes. Shaver claims a judgment of intention presupposes epistemic foreknowledge, but not the other way around [Shaver 1985]. Conversely, Bratman argues that epistemic foreknowledge combined with action must imply intention [Bratman 1990]. Acknowledging these different positions, our model makes the weaker inference that when an agent is certain of an outcome and performs or authorizes the action, it implies only some non-zero level of intention. When an agent orders an action that has multiple alternative outcomes and the performing agent is allowed to choose between them, outcome intention is entailed only for the performing agent.

The distinction between action and outcome intent applies to coercion as well. Where an imperative command to act is given by an agent in a position of authority, some amount of action coercion is inferred. It may or may not be effective – this is known only by comparison with later actions. Outcome coercion is inferred by the same logic as outcome intention: both agents must have foreknowledge of the outcome at the time of the coercion and other outcome options must not be equally available to the coerced agent. Furthermore, an agent with prior intention is not coerced by being ordered to do what he or she already intended.

Explicit communication of an expected future outcome entails attribution of some amount of foreknowledge of that outcome to the speaker and the hearer. This foreknowledge may be accurate or not. When the communicated claim is unqualified, our model infers equality to an upper limit point of certainty. We do not address the issue of deception.

These attributions are temporally bounded. An attribution holds over an interval that contains or is overlapped by the interval of the event that provided evidence for the attribution. Attributions are assumed to persist until they meet an event that provides evidence for a different value for that variable.

Judging responsibility

We represent judgments of responsibility as nonnegative continuous parameters whose values are constrained by ordinal relationships. These constraints may involve qualitative values on a totally ordered scale (e.g., some, none) as well as comparisons with other agents (e.g., agent X is more responsible than agent Y). These values and constraints are derived in our model from the attributions along the dimensions of responsibility.

[1] Social psychological research cited in [Mao 2006] indicates that there are differences in the processes used for responsibility for positive events and negative events, hence the exclusive focus on negative events here.

Causality constrains eligibility for responsibility. The agent that performed the action that caused the outcome is eligible of course. Where an agent is in a position of authority over the action that caused the outcome, that agent is also eligible. Authority is inferred based both on domain knowledge of organizational structure and the negotiation structure of the discussion. In both cases, the agent is *responsible by action.* In the case of coercion, the coercing agent is *responsible by coercion* and is also eligible for responsibility for the outcome. An agent in a position of authority who coerces is considered responsible by coercion rather than action.

Given our omission of the more special-case dimension of appreciation, Shaver's attribution process displays four distinct modes of judgment: *causal without foreknowledge*, *causal without intent*, *intentional but coerced* and *intentional in the absence of coercion*. Responsibility is strictly increasing across these modes, in that order. Within each state, responsibility is qualitatively proportional ($\propto_{Q+}$) to a different attribution variable.

We claim that the causal without foreknowledge mode is better understood as a case of achievement failure, making responsibility qualitatively proportional to causal controllability. The causal without intent and intentional but coerced modes also rely on attribution variables that impact causal controllability (foreknowledge and coercion, respectively) and thus ordering constraints cannot be placed between them and the achievement failure mode. The intentional in the absence of coercion mode, by contrast, assumes causal controllability and thus can be attributed higher responsibility.

These modes are formalized by six model fragments (views) consisting of *conditions* and *consequences*. The first two modes translate directly into two views. The third and fourth modes each translate into two views based on whether the agent being considered is responsible by action or coercion. Intention and foreknowledge are measured at the time of the action or coercion. Formal details of the views are provided in predicate calculus below. Due to space limitations, only the four views relevant to the experiment are included.

```
View: AchievementFailure-Foreknowledge
Conditions:
 responsibleByActionFor(?agent, ?action, ?outcome) ∧
 ValueDuringFn(
  ForeknowledgeFn(?agent, causes(?action, ?outcome)),
  ?action) = 0
Consequences:
 ResponsibilityFn(?agent, ?outcome) ∝Q+
  CausalControlFn(?agent, ?action, ?outcome)
```

```
View: IntentionalButCoerced
Conditions:
 responsibleByActionFor(?agent, ?action, ?outcome) ∧
 ValueDuringFn(
  IntentionFn(?agent, ?action, ?outcome),
  ?action) > 0 ∧
 ValueDuringFn(
  CoercionFn(?coercer, ?agent, ?coercion-action,
           ?action, ?outcome),
  ?action) > 0
Consequences:
  ResponsibilityFn(?agent, ?outcome) ∝Q+
   ValueDuringFn(
    CoercionFn(?coercer, ?agent, ?coercion-action,
             ?action, ?outcome),
    ?action)
```

```
View: Intentional
Conditions:
 responsibleByActionFor(?agent, ?action, ?outcome) ∧
 ValueDuringFn(
  IntentionFn(?agent, ?action, ?outcome),
  ?action) > 0 ∧
 ¬∃?coercer, ?coercion-action(
    ValueDuringFn(
     CoercionFn(?coercer, ?agent, ?coercion-action,
              ?action, ?outcome),
     ?action) > 0)
Consequences:
  ResponsibilityFn(?agent, ?outcome) ∝Q+
   ValueDuringFn(
    IntentionFn(?agent, ?action, ?outcome),
    ?action)
```

```
View: IntentionalByCoercion
Conditions:
 responsibleByCoercionFor(?agent, ?coercion-action,
                        ?action, ?outcome) ∧
ValueDuringFn(
 IntentionFn(?agent, ?action, ?outcome),
 ?coercion-action) > 0 ∧
¬∃?coercer2, ?coercion-action2(
   ValueDuringFn(
    CoercionFn(?coercer2, ?agent, ?coercion-action2,
             ?action, ?outcome),
    ?coercion-action) > 0
Consequences:
  ResponsibilityFn(?agent, ?outcome) ∝Q+
   ValueDuringFn(
    IntentionFn(?agent, ?action, ?outcome),
    ?coercion-action)
```

For each scenario with a negative outcome and some number of agents, our model infers which agents bear some level of responsibility, what mode of judgment they fall into and what qualitative proportionalities constrain their amount of responsibility. Given a number of such scenarios, our model is able to infer ordinal constraints on responsibility for pairs of agents both within and across the scenarios, although, given the qualitative nature of the constraints, total orderings may not always be possible. For situations where two responsibility judgments fall into different modes, the inference is straightforward. For judgments within the same mode, relative amounts of responsibility are inferred when ordinal relationships between the control parameters are known.

As in Mao's model, the strength of coercion is determined by evidence of intention prior to the coercing. An agent who is known to have not intended the action or outcome prior to being ordered to do it is attributed greater coercion than an agent whose prior intention is unknown. Formally, the ordinal constraint is inferred as:

```
R1: ValueDuringFn(
     CoercionFn(?coercer1, ?agent1, ?coercion-action1,
                ?action1, ?outcome1),
     ?action1) > 0
    ValueDuringFn(
     IntentionFn(?agent1, ?action1, ?outcome1),
     ?coercion-action1) = 0 ∧
    ValueDuringFn(
     CoercionFn(?coercer2, ?agent2, ?coercion-action2,
                ?action2, ?outcome2),
     ?action2) > 0 ∧
    ¬(ValueDuringFn(
       IntentionFn(?agent2, ?action2, ?outcome2),
       ?coercion-action2) = 0)
⇒ ValueDuringFn(
    CoercionFn(?coercer1, ?agent1, ?coercion-action1,
               ?action1, ?outcome1), ?action1) >
   ValueDuringFn(
    CoercionFn(?coercer2, ?agent2, ?coercion-action2,
               ?action2, ?outcome2), ?action2)
```

In the view of achievement failure with regard to foreknowledge, there is a chain of communication of incorrect information. Agents that are further down that chain from the source of the information are attributed less control than agents closer to the source. Formally, the ordinal constraint is inferred as:

```
R2: responsibleByActionFor(?acting-agent, ?action,
                           ?outcome) ∧
    ValueDuringFn(
     ForeknowledgeFn(?acting-agent,
                     ¬causes(?action, ?outcome)),
     ?action) > 0 ∧
    sourceOfForeknowledge(
     ?agent1,
     ForeknowledgeFn(?acting-agent,
                     ¬causes(?action, ?outcome))) ∧
    ¬sourceOfForeknowledge(
      ?agent2,
      ForeknowledgeFn(?acting-agent,
                      ¬causes(?action, ?outcome)))
⇒ CausalControlFn(?agent1, ?action, ?outcome) >
    CausalControlFn(?agent2, ?action, ?outcome)
```

Shaver argues that in causality, omission is just as blameworthy as commission. In our model we extend this allowance to the dimension of coercion. As stated in rule R2, an agent who is in a position of authority over a causal action is considered eligible for responsibility. If the authority is aware of a possible negative outcome from the subordinate's actions, yet does not coerce the subordinate away from that outcome, then they are guilty of abdicating authority. Under these circumstances the authority is subject to the same evaluation of intention as the underling. However, if the authority is unaware of the underling's actual intention to cause that outcome, then his or her outcome intention is constrained to be less than the intention of the underling. Formally, these inferences are:

```
R3: causes(?action, ?outcome) ∧
    performedBy(?action, ?underling) ∧
    authorizedBy(?action, ?authority) ∧
    ValueDuringFn(
     ForeknowledgeFn(?authority,
                     ¬causes(?action, ?outcome)),
     ?action) > 0 ∧
    ¬∃?coercion-action(
        ValueDuringFn(
         CoercionFn(?authority, ?underling,
                    ?coercion-action,
                    ?action, ?outcome),
         ?action) > 0)
⇒ abdicatedAuthority(?authority, ?underling,
                     ?action, ?outcome)
```

```
R4: abdicatedAuthority(?authority, ?underling,
                       ?action, ?outcome) ∧
    ¬∃?sit(
     ValueDuringFn(
      ForeknowledgeFn(
       ?authority,
       ValueDuringFn(
        IntentionFn(?underling, ?action, ?outcome),
        ?sit) > 0)
      ?sit) > 0 ∧
     overlaps(?sit, ?action)
⇒ ValueDuringFn(
    IntentionFn(?underling, ?action, ?outcome),
    ?action) >
   ValueDuringFn(
    IntentionFn(?authority, ?action, ?outcome),
    ?action)
```

Finally, the outcome intention of an agent who chooses not to coerce, even one in authority, must be considered less than that of an agent who chooses to coerce. Formally:

```
R5: responsibleByAction(?coerced,
                        ?action1, ?outcome1) ∧
    ValueDuringFn(
     CoercionFn(?agent1, ?coerced, ?coercion-action,
                ?action1, ?outcome1),
     ?action1) > 0 ∧
    abdicatedAuthority(?agent2, ?underling,
                       ?action2, ?outcome2)
⇒ ValueDuringFn(
    IntentionFn(?agent1, ?action1, ?outcome1),
    ?action1) >
   ValueDuringFn(
    IntentionFn(?agent2, ?action2, ?outcome2),
    ?action2)
```

Experiment

Mao presents an evaluation of her system against human data collected in a survey of 30 respondents. The survey presented four scenarios, variations starting with the "*company program*" scenario used by Knobe [Knobe 2003], replicated below. The scenarios involve two agents, a chairman and a vice president, and a negative outcome of environmental harm. Each scenario was followed by a set of Yes/No questions intended to validate the judgments of intermediate variables, including the attribution variables, and a final question asking the respondent to score the blame each agent deserved on a scale of 1-6. Due to space limitations, we refer the reader to [Mao 2006] for details on the data collection process.

Corporate Program Scenarios

Scenario 1. The vice president of Beta Corporation goes to the chairman of the board and requests, "Can we start a new program?" The vice president continues, "The new program will help us increase profits, and according to our investigation report, it has no harm to the environment." The chairman answers, "Very well." The vice president executes the new program. However, the environment is harmed by the new program.

Scenario 2. The chairman of Beta Corporation is discussing a new program with the vice president of the corporation. The vice president says, "The new program will help us increase profits, but according to our investigation report, it will also harm the environment." The chairman answers, "I only want to make as much profit as I can. Start the new program!" The vice president says, "Ok," and executes the new program. The environment is harmed by the new program.

Scenario 3. The chairman of Beta Corporation is discussing a new program with the vice president of the corporation. The vice president says, "The new program will help us increase profits, but according to our investigation report, it will also harm the environment. Instead, we should run an alternative program, that will gain us fewer profits than this new program, but it has no harm to the environment." The chairman answers, "I only want to make as much profit as I can. Start the new program!" The vice president says, "Ok," and executes the new program. The environment is harmed by the new program.

Scenario 4. The chairman of Beta Corporation is discussing a new program with the vice president of the corporation. The vice president says, "There are two ways to run this new program, a simple way and a complex way. Both will equally help us increase profits, but according to our investigation report, the simple way will also harm the environment." The chairman answers, "I only want to make as much profit as I can. Start the new program either way!" The vice president says, "Ok," and chooses the simple way to execute the new program. The environment is harmed.

Mao and Gratch results

	Human Data		Mao Model		
	Chair	VP	Chair	VP	Degree
Scenario1	3.00	3.73		Y	Low
Scenario2	5.63	3.77	Y		Low
Scenario3	5.63	3.23	Y		Low
Scenario4	4.13	5.20		Y	High

Table 1. Mao and Gratch results

Table 1 shows, for each scenario, the average blame attributed to each agent by the survey respondents, the single choice of the blameworthy agent made by Mao's system and the degree of responsibility for that agent asserted by Mao's system. In each scenario, Mao's model correctly selects the agent who receives the higher degree of blame, but with the incorrect implication that the other agent involved is free of responsibility. The assignments of degree of responsibility in Mao's model do not match the human data.

Our experiment

In order to reduce tailorability, we semi-automatically encoded the scenarios using our Explanation Agent natural language understanding system (EA NLU) [Kuehne & Forbus 2004]. This required extending EA NLU in several ways, including handling tense and aspect, modal statements, explicit utterances and identifying communication events in texts. EA NLU utilizes a knowledge base based on ResearchCyc[2] contents, augmented with our own representations for QP theory. Processing of sentences and constructing predicate calculus representations is automatic, but experimenters are expected to provide choices when the system constructs multiple interpretations due to ambiguities. While fully automatic processing would be preferable, given the state of the art in NLP, that is impractical. Using an NLU system and off-the-shelf knowledge base contents greatly reduces the degree of tailorability and simplifies stimulus construction.

Qualitative model results

Figure 1 shows the ordinal constraints inferred by our model on the amount of responsibility for the agents across all four scenarios, together with labels indicating the average blame attributed to each by the survey respondents.

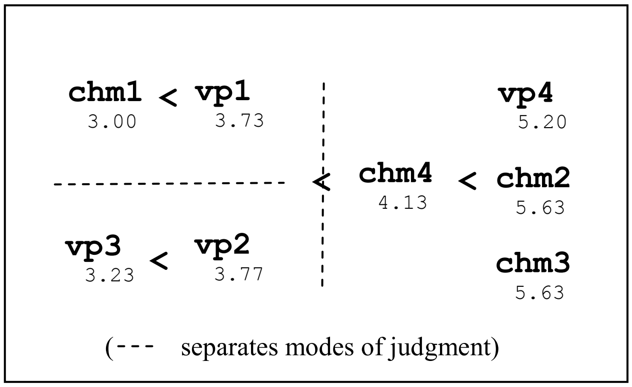

Figure 1. Ordinal constraints on responsibility and average participant attribution numbers

The eight agents being considered fall into three of the four modes of judgment. By the ordering constraints on the modes, all agents in the intentional mode of judgment are attributed more responsibility than the agents in the other two. The chairman and vice president in scenario 1 fall into the `AchievementFailure-Foreknowledge` view. Within this view, the responsibility of each agent is qualitatively proportional to the amount of causal control each agent is judged to have over the outcome. The chairman is judged to have less control by rule R2 and thus less responsibility. The vice president in scenario 2 and the vice president in scenario 3 fall into the `IntentionalButCoerced` view. Their respective degree of responsibility is qualitatively proportional to the amount of applied coercion. There is no indication of the outcome intention of the vice president in scenario 2 prior to the

[2] www.cycorp.com

coercion action, while the vice president in scenario 3 clearly shows lack of outcome intention prior to being coerced. The vice president in scenario 3 is therefore judged to have a higher degree of coercion by rule R1 and thus a lower degree of responsibility. The chairman and vice president in scenario 4 fall into the `Intentional` view while the chairmen from scenarios 2 and 3 fall into the `IntentionalByCoercion` view. Responsibility for all three is qualitatively proportional to their outcome intention. The chairman in scenario 4 abdicated authority to the vice president, as captured by rule R3. But since the outcome was not coerced and there was no prior knowledge of the vice president's intention, he is constrained to have a lower degree of intention than the other three by rules R4 and R5. This results in lower responsibility, while the other three remain unordered.

In 21 of the 28 possible comparisons between agents our model infers which agent should receive more blame. All of these 21 comparisons match the results from the human respondents. In 3 of the 7 remaining comparisons, our system establishes a constraint between the degree of responsibility and the value of an attribution variable for each agent, but cannot infer an ordinal relation between the control variables. In the remaining 4 cases, comparing agents in the achievement failure mode to agents in the intentional but coerced mode, the interaction between the control variables is undefined. The impact of the perception of coercion on the perception of causal control cannot be compared to the impact of achievement concerns such as ability and effort. We suspect that the effort displayed by the vice president in scenario 3 contrasts with the clear lack of effort by the vice president in scenario 2 and a presumed lack of effort on behalf of the vice president in scenario 1. Nevertheless, this interaction is not accounted for by the current model.

Based on the 3 cases where our model infers a constraint with a free variable, we can make predictions about additional constraints in the attribution variables. Given that the respondents attributed equal blame to the chairmen in scenarios 2 and 3, our model predicts that they would judge the outcome intention of the chairmen as being equal as well. This is consistent with the implicit claim in attribution theory that, while coercion mitigates the responsibility of the coerced, it has no such effect on the responsibility of the coercer. Finally, since respondents attributed less blame to the vice president in scenario 4 than to the chairmen in scenarios 2 and 3, our model predicts that they would judge the outcome intention of that vice president to be less than the outcome intention of either chairman.

Conclusion and Future Work

We have shown that QP theory can be used to formally encode a model for attributing responsibility for negative outcomes, based on attribution theory. Our model explains the corporate scenario data better than Mao's model does, due to our use of qualitative representations instead of categorical, Boolean values. While a purely qualitative model would not be sufficient for all purposes – for example, deciding whether or not someone was blameworthy enough to report an action – our evaluation suggests that qualitative modeling captures an important level of reasoning about social situations.

This work represents part of a larger effort to model and reason about culturally-sensitive moral decision-making. In that context, we are expanding the capabilities of the EA NLU system to capture a broader range of narratives about real world situations. By utilizing natural language we reduce the tailorability of our representations and increase the contextual details encoded for each scenario. We plan to expand the factors that go into making attribution judgments beyond simple action-outcome sequences and order negotiation speech acts. In doing so we will be able to further evaluate of the validity of those judgments and the predictions made by this model regarding the attribution of blame.

Acknowledgements

This research was supported by the Air Force Office of Scientific Research. We thank Jon Gratch for useful conversations.

References

Bratman, M. 1990. What is intention? In P. Cohen, J. Morgan & M. Pollack eds., *Intentions in Communication*. MIT Press.

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

Forbus, Kenneth D. & Sven Kuehne. 2005. Towards a qualitative model of everyday political reasoning. *Proceedings of the Nineteenth International Qualitative Reasoning Workshop*.

Heider, F. 1958. *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc.

Kamps, J. and Peli, G. 1995. Qualitative reasoning beyond the physics domain: The density dependency theory of organizational ecology. *Proceedings of the Ninth International Qualitative Reasoning Workshop.*

J. Knobe. 2003. Intentional Action and Side-Effects in Ordinary Language. *Analysis*, 63:190-193.

Kuehne, S. and Forbus, K. (2004). Capturing QP-relevant information from natural language text. *Proceedings of QR04*, Evanston, Illinois, August..

Mao, W. 2006. *Modeling Social Causality and Social Judgment in Multi-Agent Interactions.* Doctoral dissertation, University of Southern California, Los Angeles, California.

Mao, W. and Gratch, J. 2005. Social Causality and Responsibility: Modeling and Evaluation. *Fifth International Conference on Interactive Virtual Agents*.

K. G. Shaver. 1985. *The Attribution Theory of Blame: Causality, Responsibility and Blameworthiness.* Springer-Verlag.

Steinmann, C. 1997. Qualitative reasoning on economic models. *Proceedings of the Eleventh International Qualitative Reasoning Workshop.*

B. Weiner. 1995. *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. Guilford Press.

B. Weiner. 2001. Responsibility for Social Transgressions: An Attributional Analysis. In B. F. Malle, L. J. Moses and D. A. Baldwin eds. *Intentions and Intentionality: Foundations of Social Cognition*, pp. 331-344. MIT Press.