A CBR Knowledge Representation
for Practical Ethics

Kevin D. Ashley and Bruce M. McLaren
University of Pittsburgh
Intelligent Systems Program, School of Law,
and Learning Research and Development
Pittsburgh, Pennsylvania 15260
(email: ashley@vms.cis.pitt.edu, bmm@cgi.com)

Abstract

TRUTH-TELLER, a program for testing a Case-Based Reasoning (CBR) knowledge representation in practical ethics, compares cases presenting ethical dilemmas about whether to tell the truth. Its comparisons list ethically relevant similarities and differences (i.e., reasons for telling or not telling the truth which apply to both cases, and reasons which apply more strongly in one case than another or which apply only to one case). The reasons may invoke ethical principles or selfish considerations. We describe a knowledge representation for this practical ethical domain including representations for reasons and principles, truth telling episodes, contextually important scenarios, and comparison rules. In a preliminary evaluation, a professional ethicist scored the program's output for randomly-selected pairs of cases. The work contributes to AI CBR efforts to integrate general principles and context-sensitive information in symbolically assessing case similarity and to model comparing problems to paradigmatic cases. It also furthers research on cognitive and philosophical models of ethical judgment and decision-making.

1. Introduction

A primary goal for AI CBR research is to identify ways that human reasoners employ cases to evaluate problems comparatively. In a variety of professional domains and in "common sense" reasoning, humans employ techniques to draw inferences about problem situations by comparing them to past cases. Case-based comparative evaluation skills appear to help human reasoners to deal with weak analytic domain/task models. Such models are too weak to support constructing proofs of the correct answers to problems. Nevertheless, the models do support constructing arguments comparing the problems to past cases and drawing useful conclusions. We will refer to them as comparative evaluation models.

Practical ethical reasoning is a domain in which a comparative evaluation model supplements a weak analytic model. Although philosophers have explored a variety of techniques for solving practical dilemmas by resolving conflicting ethical principles, the attempts have largely failed. Deductive reasoning does not work, because ethical principles are often inconsistent and their antecedents are not well defined. "No moral philosopher has ever been able to present a system of moral rules free of these kinds of conflicts between principles and exceptions to principles" [Beauchamp and McCullough, 1984, p. 16]. If one could assign weights to competing principles, resolving them would simply be a matter of comparing the weights. However, ``the metaphor of the `weight' of a principle of duty has not proven amenable to precise analysis'' [Beauchamp and McCullough, 1984, p. 16]. More recently, ethicists have proposed alternative case-based (i.e., "casuistic") models in which problems are systematically compared to past or paradigmatic cases that bear on the decision [Strong, 1988, Jonsen and Toulmin, 1988, Schaffner, 1990].

Carson Strong, for instance, has proposed a systematic, five-step Case Comparison Method for justifying decisions in practical ethics. The last three steps involve comparing the problem with relevant paradigmatic cases. The comparison aims to illuminate the factors that favor applying or not applying the principles used in the

paradigm [Strong, 1988]. Strong's model will strike the AI CBR researcher intuitively as nearly computational:

1. Identify middle-level principles and role-specific duties pertinent to situation.

2. Identify alternative courses of action that could be taken.

3. Identify morally relevant ways in which cases of this type can differ from one another (i.e., factors). Comparing with other cases of the same type also helps identify factors.

4. For each option, identify paradigm case in which option would be justifiable. Paradigms can be actual or hypothetical cases. Identify middle-level principle which would provide that justification.

5. Compare case at hand with paradigm cases. Determine which paradigms it is "closest to" in terms of presence of morally relevant factors [Strong, 1988].

Computationally realizing a model like Strong's is interesting from an AI CBR viewpoint because, while it is similar to various AI CBR models (e.g., HYPO in comparing cases along factors [Ashley, 1990], PROTOS in comparing problems to prototypical or paradigmatic cases [Bareiss, 1989], Veloso's PRODIGY/ANALOGY in comparing cases to select appropriate actions to apply [Veloso, 1992]), it introduces important components of CBR that have not yet been modeled such as: (1) symbolically comparing problems and paradigmatic cases to resolve conflicts among applicable general principles and (2) adequately accounting for a problem's specific contextual circumstances in deciding how to resolve conflicting general principles.

As a first step toward computationally realizing a model like Strong's, we have designed and built a computer program, TRUTH-TELLER, whose task is to draw ethically relevant comparisons between practical ethical problems in the truth telling domain. We have focused on case comparison initially because any casuistic model requires a capacity for symbolic case comparison (see e.g., the last three steps of Strong's process). Like [Edelson, 1992], we represent principles at various levels of abstraction which apply to cases, but our representation is designed for comparing cases in ways other than in terms of the principles' abstraction levels. High level case information (i.e., legal theories) and prototypical cases have been integrated into case retrieval and argumentation in [Rissland, Skalak, and Friedman, 1993]. We attempt, however, to elaborate a more comprehensive knowledge representation for a more reflective comparative justification with principles, cases, and actions in a context where cases regularly have more than two possible outcomes. Unlike [Kass, Leake, & Owens, 1986, Edelson, 1992], we do not attempt a narrative representation of the truth telling episodes in terms of either the protagonists' goals and plans or expectation violations.

2. The Comparative Evaluation Model in TRUTH-TELLER

TRUTH-TELLER (TT), a program for testing and developing a CBR knowledge representation in practical ethics, compares cases presenting ethical dilemmas about whether to tell the truth. Its comparisons list ethically relevant similarities and differences (i.e., reasons for telling the truth or not telling the truth which apply in both

cases, and reasons which apply more strongly in one case than another or which apply only to one case). The reasons may invoke ethical principles or selfish considerations. The knowledge representation for this practical ethical domain includes representations for reasons and principles, truth telling episodes, contextually important scenarios, and comparison rules. Our ultimate goal is to see whether a comparative evaluation model like TRUTH-TELLER's could help drive a tutorial program in practical ethical reasoning. Currently, we are recording protocols of high school students' arguments about the same ethical dilemmas contained in TT's Case Knowledge Base.

Currently, TRUTH-TELLER has 23 cases adapted from a game called Scruples (TM). Two of those cases, Rick's case and Wanda's case, are shown at the top of Figure 1, followed by the program-generated comparison of the cases. After reciting the cases TT lists ethically relevant similarities and differences between the cases, differences it finds or infers using five knowledge sources:

- **Truth telling episodes** including for each episode: (a) the actors (i.e. the truth teller, truth receiver, others affected by the decision), (b) relationships among the actors (e.g. familial, seller-customer, attorney-client, student-teacher), (c) the truth teller's possible actions (i.e. telling the truth, silence, telling a lie, or taking some alternative action), and (d) reasons for and against each possible action. The information is represented in semantic networks using the knowledge representation language, LOOM [MacGregor, 1991] [Woods & Schmolze, 1990].

- **Relations Hierarchy**: The relationships among the actors are drawn from the Relations Hierarchy, a taxonomy of approximately 80 possible relationships among the participants in a truth telling episode. Mid-level relationships include familial, commercial, and acquaintance relations. Higher level relationships include minimal-trust and high-trust relations and authority relations. The Relations Hierarchy is used to infer which relationships are "similar" for purposes of identifying the levels of trust and responsibility that apply among the participants (i.e., the applicable scenarios. See below).

- **Reason Hierarchy**: A reason is a rationale for taking an action. We represent reasons as a hierarchy of concepts, the Reason Hierarchy. Reasons have four facets: type, criticality, altruistic?, and principled?; each facet is important to ethical decision-making and is represented as a distinct branch of the Reason Hierarchy. The Reason Hierarchy is used to characterize abstractly the reasons for and against an action according to these facets. Based on Bok's formulation in (Bok, 1989), a reason's type is based on four underlying general principles for telling or not telling the truth: fairness, veracity, beneficence, and nonmaleficence. We also represent a variety of more specific principles.

- **Scenario Hierarchy**: In determining whether or not to tell the truth, contextual information is important. One needs to consider such things as the consequences of an action, the reasonable expectations of truthfulness that apply in different social contexts, and the level of trust or reliance among the actors. We have identified approximately 15 types of contextual information, we call them truth-telling scenarios, and have organized them into a Scenario Hierarchy. Our scenarios include context-specific considerations such as: Is there a relationship of

authority between the teller and receiver? of trust? Are others affected by the decision to tell the truth? Is the action a matter of telling an out-and-out lie or simply keeping silent? If the action is telling a lie, is it premeditated or not? What is the nature of and how severe are the consequences of the action? Are there alternative actions that obviate the need to tell the lie or disclose the information? Are the participants involved in a game or activity governed by disclosure rules?

- **Comparison rules**: We have defined 58 Comparison Rules for deciding whether one case presents a stronger or weaker justification than another case for taking a course of action such as disclosing information or not telling a lie. From the information contained in the cases' applicable scenarios, actions, reasons, and principles, the rules infer relevant ethical similarities and differences between the cases. The rules' left hand sides employ classified scenarios, actions, reasons, and reason-associated principles. The right hand sides make assertions about the relative strength in the two cases of the conclusions that the teller should tell the truth or disclose information (or not).

The goal of TRUTH-TELLER's knowledge representation design is to enable the program to make context sensitive ethical comparisons of cases. To this end, we have designed a knowledge representation that enables TT, in comparing cases, to go some way beyond matching the reasons listed in the two cases' representations. In effect, TT reasons about reasons in the context of a particular pair of cases in a variety of ways. We will illustrate the effect of this reasoning in an extended example comparing the output text of Figure 1 with the initial lists of reasons in the semantic networks for Rick's and Wanda's cases, shown in Figure 2. Specifically, the example illustrates how TRUTH-TELLER: (1) classifies reasons as principled, self-motivated, or altruistic, (2) elicits the principles underlying the reasons, (3) matches reasons to find shared reasons for an action (similarities) and unshared reasons or reasons that apply more strongly in one case than another (differences). In the example, the program tailors the comparison based on (4) differences in criticality of the reasons, (5) differences in the participants' roles, (6) similarity in types of reasons considered in the aggregate, (7) similarities in the presence of untried alternatives and of (8) others affected by an action.

"Rick's father is having an affair. Rick's mother is unaware of it. Should Rick tell his mother?"

"Wanda's brother, Paul, is a real playboy. He's setting up a young woman, whom Wanda knows casually, for a big disappointment. The young woman asks Wanda if Paul loves her. Should Wanda tell her the truth?"

SIMILARITIES:

R17

Rick and Wanda have a common principled reason for telling the truth, i.e. providing the right to disclosure of information. This reason is supported by the ethical principle: 'When one knows that another person is relying on the disclosure of information, that information should be disclosed.'

Rick and Wanda have a common principled reason for not telling the truth, i.e. avoiding becoming a 'tattle tale'. This reason is supported by the ethical principle: 'One should not 'tattle' on another person.'

R25

Rick and Wanda have purely altruistic reasons for not telling the truth. To some extent this would excuse each for not telling the truth.

Both of the cases involve someone other than the person who must decide whether or not to tell the truth and the person who will hear the truth. Ricks-father is affected by Rick's decision, and Paul is affected by Wanda's decision.

Both Rick and Wanda have an alternative action that could be taken before making the truthtelling decision. Rick could approach Ricks-father before deciding whether to tell the truth or not, while Wanda could approach Paul before deciding whether to tell the truth or not.

DIFFERENCES:

The stakes for Rick are much higher than for Wanda. Rick's decision is highly important.

Rick has a reason for telling the truth -- providing fairness for a misdeed that has been committed -- that is not shared by Wanda. This reason is supported by the ethical principle: 'When one has committed a misdeed, they should be held accountable to those who have been affected.'

Rick has a reason for not telling the truth that is not shared by Wanda. The reason is avoiding major emotional distress. The ethical principle: 'One should protect themself and others from serious emotional distress.' supports this reason.

Wanda has a reason for not telling the truth that is not shared by Rick. The reason is protecting against a possibly false accusation. The ethical principle: 'One should not make accusations unless they are sure their accusations are well-founded.' supports this reason.

Wanda has a second reason for not telling the truth, i.e. avoiding minor emotional distress, that is not shared by Rick. However, this reason is not supported by any ethical principle.

Ricks-mother has authority over Rick, but Young-woman and Wanda are on a relatively equal basis. Since Rick is subject to this authority, it may increase the pressure to tell the truth.

R36

Rick has a high level of duty to tell the truth to Ricks-mother. However, the duty Wanda has to Young-woman is much lower.

Ricks-mother is likely to have great trust in Rick telling the truth. However, the trustYoung-woman has in Wanda is much lower

Wanda is confronted with telling an outright lie, while Rick must decide whether to remain silent about the truth. This tends to make Wanda's decision more difficult, i.e. it is typically less excusable to lie than to remain silent about the truth.

Figure 1: TRUTH-TELLER's Output Comparing Rick's and Wanda's Cases

3. An Extended Example of Case Comparison

TRUTH-TELLER applies its five knowledge sources in a two-step process of classification and comparison:

Classification Step: For each of two selected input cases, classify the case's manually-constructed semantic representation by (1) all applicable role-specific scenarios, (2) all applicable actions, and (3) an abstract characterization of the reasons for those actions (using the Reason Hierarchy).

Comparison Step: Attempt to apply all Comparison Rules to the classified cases. The output of this step is a list of relevant similarities and differences which is translated into a comparison text (using the Scenario and Reason Hierarchies).

TRUTH-TELLER's process starts with semantic representations of each of the cases. The representation of a case is an interpretation of its language and is filled in manually. Figure 3 depicts the semantic representation of Rick's case. Rick is the truth teller (i.e., it is he who is confronted with the decision to tell the truth or not.) Rick's mother will hear the truth, should Rick decide to divulge it, and thus is the truth receiver. Finally, Rick's father is an affected other, since he is the subject of any truth-telling disclosure, and he would be affected by the disclosure.

The comparison starts when TRUTH-TELLER accepts a semantic representation of each of the cases. The representation of a case is an interpretation of its language and is filled in manually. Figure 2 depicts the semantic representation of Rick's case. Rick is the truth teller (i.e., it is he who is confronted with the decision to tell the truth or not.) Rick's mother will hear the truth, should Rick decide to divulge it, and thus is the truth receiver. Finally, Rick's father is an affected other, since he is the subject of any truth-telling disclosure, and he would be affected by the disclosure. The relevant roles and relationships between actors in the case are also included in the semantic representation. Some relationships and roles are provided as input (e.g., Rick is the son-of Rick's mother and father) while others are deduced by forward chaining rules (e.g., Rick's mother has-husband Rick's father, since they share a common child).

The semantic representation also contains a set of possible actions that the truth teller could take and reasons supporting each of the actions. One of the possible actions is always to tell the truth and another is some version of not telling the truth, for instance, telling a lie or keeping silent (i.e., not disclosing information). In Rick's case, the choice is between telling the truth about his father's affair or keeping silent. Since the case does not state that Rick was asked whether his father was having an affair, Rick is not confronted with telling an outright lie. Rick also has an alternative action he could take before deciding whether or not to talk with his mother; he could first speak with his father. Actions are supported by reasons; a reason is a rationale for taking an action. For example, a rationale for Rick's telling the truth is to protect his mother's right to the disclosure of information important to her. A rationale for keeping silent is to avoid inflicting serious emotional distress upon his mother.

Given the input representations and after inferring roles and relationships, TRUTH-TELLER performs the Classification Step; it classifies the cases and case components including the actions and reasons. For purposes of comparing cases, the two most critical classifications are assigning the role-specific scenarios and characterizing the reasons.
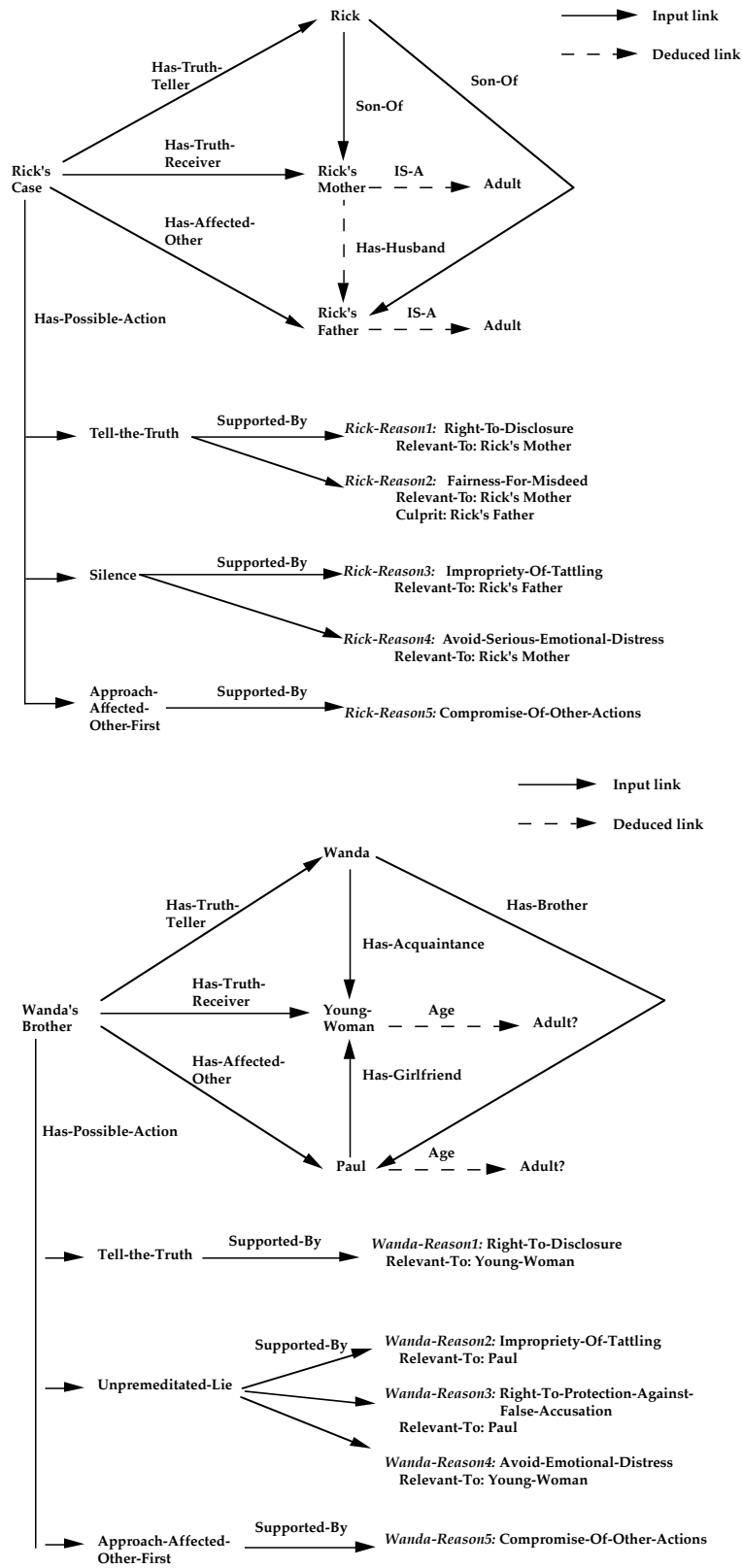
Rick

Has-Truth-Teller

Son-Of

Son-Of

Input link

Deduced link

Rick's Case

Has-Truth-Receiver

Rick's Mother — IS-A — → Adult

Has-Affected-Other

Has-Husband

Has-Possible-Action

Rick's Father — IS-A — → Adult

Tell-the-Truth — Supported-By — *Rick-Reason1:* Right-To-Disclosure
Relevant-To: Rick's Mother

Rick-Reason2: Fairness-For-Misdeed
Relevant-To: Rick's Mother
Culprit: Rick's Father

Silence — Supported-By — *Rick-Reason3:* Impropriety-Of-Tattling
Relevant-To: Rick's Father

Rick-Reason4: Avoid-Serious-Emotional-Distress
Relevant-To: Rick's Mother

Approach-Affected-Other-First — Supported-By — *Rick-Reason5:* Compromise-Of-Other-Actions

---

Input link

Deduced link

Wanda

Has-Truth-Teller

Has-Brother

Has-Acquaintance

Wanda's Brother

Has-Truth-Receiver

Young-Woman — Age — → Adult?

Has-Affected-Other

Has-Girlfriend

Has-Possible-Action

Paul — Age — → Adult?

Tell-the-Truth — Supported-By — *Wanda-Reason1:* Right-To-Disclosure
Relevant-To: Young-Woman

Unpremeditated-Lie — Supported-By — *Wanda-Reason2:* Impropriety-Of-Tattling
Relevant-To: Paul

Wanda-Reason3: Right-To-Protection-Against-False-Accusation
Relevant-To: Paul

Wanda-Reason4: Avoid-Emotional-Distress
Relevant-To: Young-Woman

Approach-Affected-Other-First — Supported-By — *Wanda-Reason5:* Compromise-Of-Other-Actions

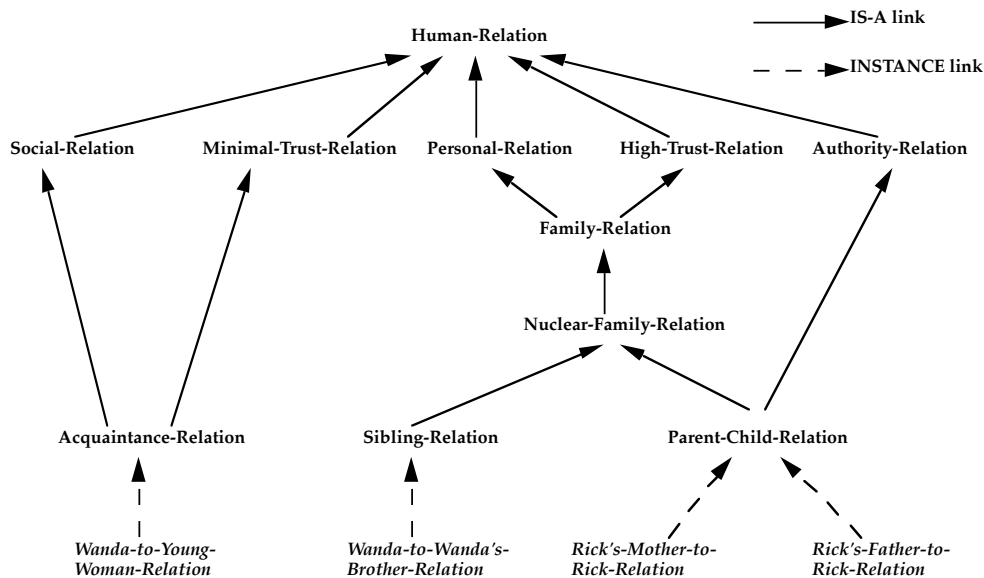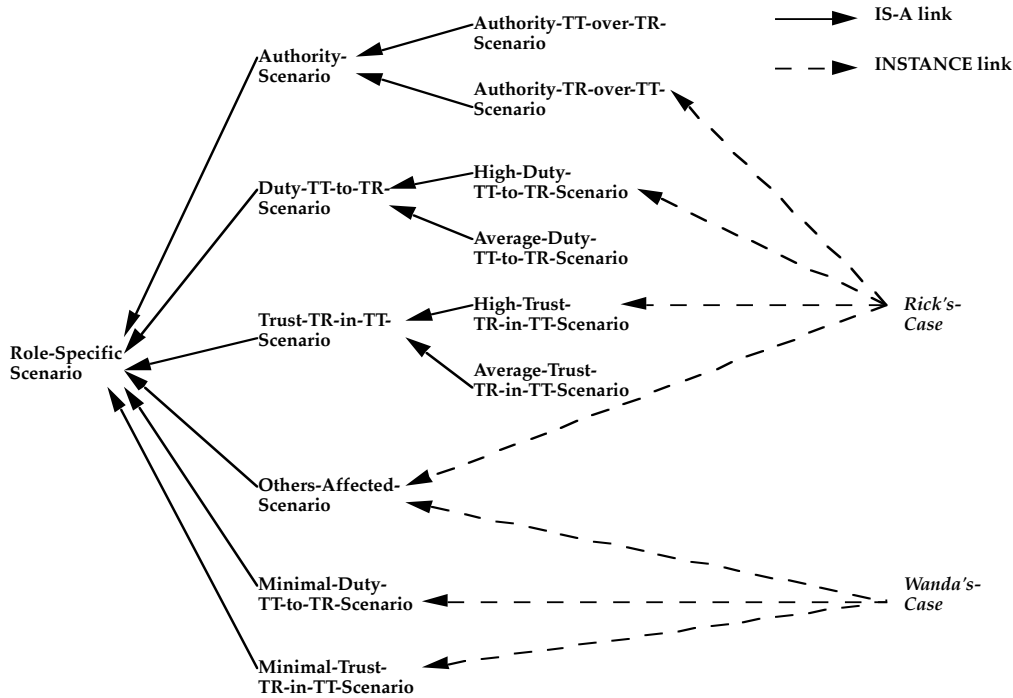**Figure 2: Semantic Network Representation of Rick's Case and Wanda's Case**

**Figure 3: Excerpts from Scenario Hierarchy and
Relations Hierarchy**

A case may be classified under any number of scenarios. LOOM's automatic classifier performs the scenario classification. Figure 3 shows a portion of the role-specific scenario hierarchy (on the top) with the Rick and Wanda cases classified. The relationships between actors in the case are the key ingredients for scenario classifications. All of the relationships in the truth-telling cases are classified within the Relations Hierarchy (a small portion of which is shown in Figure 3 on the bottom). The

Relations Hierarchy represents various types of relationships (i.e., familial, commercial, etc.) as well as important abstract information (i.e. expected level of trust, duty, and authority between the actors). The specific relations in the semantic representation of a case (e.g. son-of and has-husband as in Rick's case) are found at the lowest level of the hierarchy. Scenarios are defined with respect to relation abstractions (e.g., high trust, authority, etc.) and relation directionality (e.g., truth-teller to truth-receiver, truth-teller to affected-other, etc.) For instance, Rick's case is classified as a high trust scenario due to the relationship between Rick and his mother. Their parent-child relationship is defined, through a series of links, to be a high trust relation. By contrast, Wanda's case is classified as a minimal trust scenario, since the acquaintance relationship between Wanda and the young woman is a minimal trust relationship. Also, notice how Rick's case is classified as an authority scenario due to the parent-child relationship being an authority relationship. The cases do share one relevant scenario, however. They both involve an affected other, Rick's father and Wanda's brother.
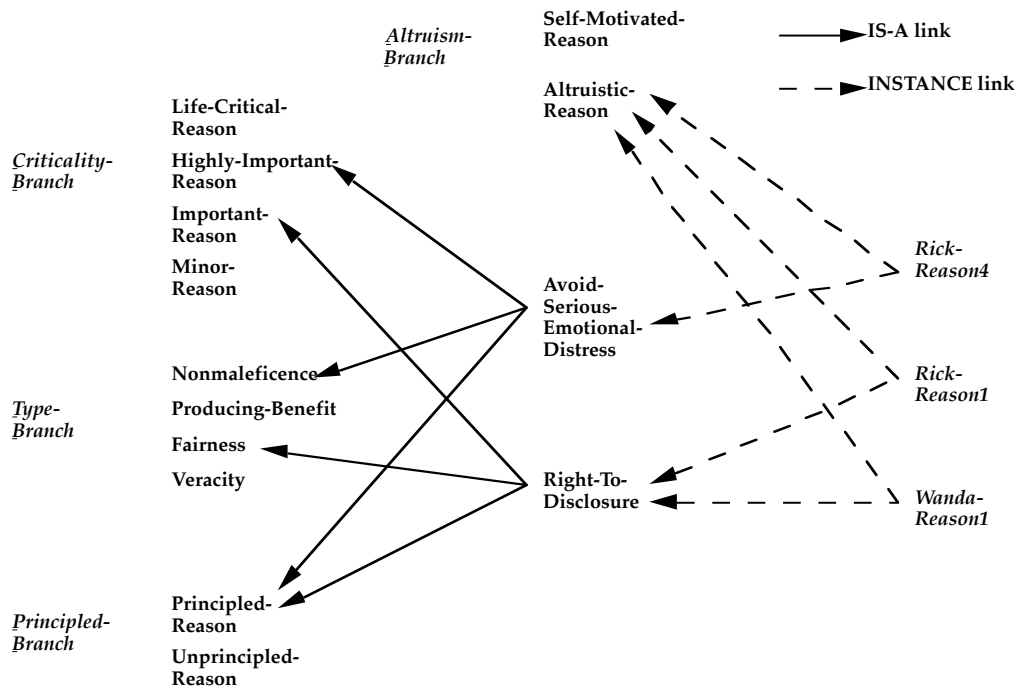
Altruism-Branch Self-Motivated-Reason Altruistic-Reason
Criticality-Branch Life-Critical-Reason Highly-Important-Reason Important-Reason Minor-Reason
Avoid-Serious-Emotional-Distress
Type-Branch Nonmaleficence Producing-Benefit Fairness Veracity
Right-To-Disclosure
Principled-Branch Principled-Reason Unprincipled-Reason
IS-A link   INSTANCE link
Rick-Reason4   Rick-Reason1   Wanda-Reason1

Figure 4: Excerpts from Reasons Hierarchy

TRUTH-TELLER also classifies the reasons associated with cases. Figure 4 shows a small portion of the Reason Hierarchy with two of Rick's reasons and one of Wanda's depicted. The top reason in the diagram -- Rick-Reason4 -- supports Rick's remaining silent (refer to Figure 2). This reason deals with Rick's sparing his mother emotional distress, should she hear of her husband's extra-marital activities. Using classification information contained in the Reason Hierarchy, the program classifies Rick-Reason4 in four ways, according to: (1) Type: Rick-Reason4 is a nonmaleficence type (i.e., intended to avoid harm). (2) Principled or not: Since this particular reason has an associated ethical principle (i.e., one should protect others from serious emotional distress) the reason is also classified as principled. (3) Criticality: qualitatively, an avoid-serious-emotional-distress reason is deemed to be highly important. (4) Altruistic or not: Rick-

Reason4 is altruistic, since it is to the benefit of Rick's mother, not to Rick himself. The other two reasons in the diagram show that Rick and Wanda share a rationale for telling the truth, namely to preserve the right of the truth receiver to hear important information. This reason is related to fairness, it is principled, and it is considered less critical than Rick-Reason4, although still considered important. Again, the reason is altruistic, as it is to the benefit of the truth receiver in both instances.

TRUTH-TELLER's comparison step attempts to apply the Comparison Rules to the cases as classified in order to draw inferences about their relevant similarities and differences and generate the comparison text as in Figure 1. We will focus on three rules that generated the text circled in Figure 1: Rules 17, 25 and 36. (The LOOM form of the rules is paraphrased for readability.) Rule 17 says, "IF CASE-1 and CASE-2 have a common principled reason for telling the truth THEN they are similar re telling truth." Here, the two cases shared the principled reason that the truth receivers had a right to the disclosures as discussed above and depicted in Figure 4. Rule-25 says, "IF CASE-1 and CASE-2 have only altruistic reasons for not telling the truth THEN they are similar re not telling truth." Since Rick and Wanda have solely altruistic reasons for not telling the truth, they both have a stronger justification for taking this action. This is an example of a comparison rule that abstracts from individual classifications and views the reasons supporting an action in the aggregate. Rule-36 says, "IF CASE-1 is a high duty scenario and CASE-2 is not THEN they are different; CASE-1 has a stronger reason to tell truth." It employs the Scenario Hierarchy (see Figure 3) to identify this important distinction between the cases. The duty between Rick and his mother is much higher than between Wanda and the young woman; this is shown by the classification of the cases within the Scenario Hierarchy. This point is important as it indicates that Rick's duty to tell the truth is higher than Wanda's. Other rules pick out the differences in criticality of consequences and similarities in the presence of untried alternatives and of others affected by the actions.

To summarize, the extended example illustrates how TRUTH-TELLER reasons about the reasons that apply in each case's particular facts and draws inferences that reflect the ethically significant differences implicit in the cases' facts. The example illustrates how TT classifies reasons as principled, self-motivated or altruistic, elicits the principles underlying the reasons, matches reasons to find shared reasons for an action (similarities) and unshared reasons or reasons that apply more strongly in one case than another (differences). The latter differences may be related to differing criticalness, differences in the participant's roles, differences in the aggregate types of reasons, and the presence of untried alternatives or of others affected by an action.

4. The Evaluation
We designed TRUTH-TELLER to reason about reasons so that it could tailor its comparisons to the particular context of the cases compared. Its ability to point out differences based on unshared compelling reasons not to tell the truth, life critical consequences of an action, varying levels of duty to tell the truth associated with particular roles and relationships among the participants, untried alternative actions and the existence of affected others enable it to make context sensitive comparisons. We undertook a preliminary evaluation to test how robustly the program could compare cases and how well the program's outputs took context into account.

Of the 23 cases in TRUTH-TELLER's CKB, thirteen initial cases were used to develop the program. The thirteen cases were employed in an initial series of interviews in which a graduate student studying medical ethics and the first author were asked to analyze, compare and contrast the cases. This information formed the basis of the knowledge representation. The remaining ten subsequent cases were added after the program had been designed and the knowledge representation had become settled.

In a preliminary experiment, we submitted fifteen pairs of cases and comparison texts (like the one in Figure 1) to an expert on moral philosophy. Thirteen of the comparison texts were generated by the TRUTH-TELLER program. The pairs were drawn from five categories. See the Table in Figure 5 for a description of the categories. Two of the comparison texts were extracted from the original interviews with the ethics graduate student (pair no. 3) and the first author (pair no. 7) and formatted and edited to look like the other comparison texts. (These pairs are marked with an "H" in the table.) The pairs of cases for the thirteen program-generated texts were selected as follows: five pairs of cases selected at random from the ten subsequent cases (category T4), five pairs of cases selected randomly consisting of one case from the initial set and one from the subsequent set (category T3), two pairs of very similar cases selected by us from the initial thirteen (category T1) and one pair of clearly distinguishable cases selected by us, one from the initial set and one from the subsequent set (category T2). The expert was not informed which texts were generated by humans and which by computer program, but he did know that some texts were generated by computer.

Of the thirteen pairs of cases for which the program generated texts, we selected three pairs and asked the expert "briefly and in writing to compare the cases in each pair from an ethical viewpoint", listing the ethically relevant similarities and differences as defined above. This task was performed before the expert saw any of the fifteen comparison texts. Then we asked the expert to "evaluate the [fifteen] comparisons as you might evaluate short answers written by your students." We asked him to assign to each text three grades on a scale of 1 to 10 for reasonableness (R score: 10 = Very reasonable, sophisticated; 1 = Totally unreasonable, wrong-headed), completeness (C score: 10 = Comprehensive and deep; 1 = Totally inadequate and shallow), and context sensitivity (CS score: 10 = Very sensitive to context, perceptive; 1 = Very insensitive to context). For each point of similarity or difference in the comparison text, we asked him also to mark the point as follows: "check" if the point is reasonable and relevant, "check+" if the point is especially good or perceptive, "check-" if the point is irrelevant, and "X" if the point is plain wrong.

The results are presented in tabular form in Figure 5. The grader treated only one text as a perfect ten, the one prepared by the medical ethics graduate student (suggesting, arguably, that context sensitive case comparison is a learned expert skill). The program generated text scores for reasonableness ranged from a high of nine to a low of six. The completeness scores ranged from a high of nine to a low of five. The scores for context sensitivity were lower, ranging from eight to two and averaging 5.3. The comparison text shown in Figure 1, for instance, which was the subject of the extended example (regarding pair no. ten, Rick's case and Wanda's case) was judged by the expert as one of the poorer comparisons in terms of context sensitivity. Interestingly, the expert graded a number of program-generated texts higher than or nearly the same as the text generated by the first author. As to the 170 points of comparison which the

program drew in total for the thirteen pairs, the expert regarded 64% as either reasonable and relevant or especially good or perceptive, 14% as irrelevant, and 22% as plain wrong.

Pair No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Avg*
Category	T1	T4	H	T2	T4	T3	H	T3	T4	T1	T4	T3	T3	T3	T4	
R Score	8	8	10	9	6	7	8	8	7	7	8	7	7	8	9	7.6
C Score	9	8	10	9	8	5	8	7	9	6	7	5	7	8	8	7.4
CS Score	8	8	10	8	4	2	7	3	6	3	7	4	3	7	6	5.3
																%**
check's	6	7	3	7	3	8	3	5	7	9	4	9	7	7	8	51
check+'s	1	3	4	1	2	2	1	2	1	1	3	1	0	3	2	13
check-'s	2	3	1	1	4	2	0	3	1	1	2	1	1	3	0	14
X's	2	2	0	1	3	4	2	5	3	4	2	4	5	1	1	22
Tot. Sim & Diff.	11	15	8	10	12	16	6	15	12	15	11	15	13	14	11	

Categories Key
T1: TRUTH-TELLER output; similar cases picked by us from initial 13
T2: TRUTH-TELLER output; clearly distinguishable cases picked by us,1 from initial 13, 1 from subsequent 10
T3: TRUTH-TELLER output; randomly chosen, 1 from initial13, 1 from subsequent 10
H: Prepared by humans

    * Average score computation excludes pairs 3 and 7 (minus humans' output)
    ** % of total number of similarities and differences generated by TRUTH-TELLER (170)

Score Key
R score (Reasonableness):
    10 = Very reasonable, sophisticated;
    1 = Totally unreasonable, wrong-headed
C score (Completeness):
    10 = Comprehensive and deep;
    1 = Totally inadequate and shallow
CS score (Context Sensitivity)
    10 = Very sensitive to context, perceptive
    1 = Very insensitive to context

Key re Individual Similarities & Differences
check = point is reasonable and relevant
check+ = point is especially good or perceptive
check- = point is irrelevant
X = point is plain wrong

Figure 5: Evaluation Table

5. Discussion and Conclusions

The evaluation suggests that the program displays a capacity for intelligently comparing truth-telling episodes. The knowledge representation was general enough to enable TRUTH-TELLER to draw reasonable comparisons of randomly selected pairs of cases from beyond the initial set used to build the representation.

The knowledge representation also was robust enough to enable comparison of the same cases in different contexts. Seven of the cases were used in more than one comparison. The program was able to draw a comparison in each context in which those cases appeared. The contexts must have been fairly different because the expert expressed difficulty working through the some of the comparisons, "perhaps because the focus is on comparison, and the cases kept appearing in new comparison contexts."

While the program's context sensitivity scores were lower than the other scores, three of its CS scores were higher than and two tied one of the human's CS scores. Since being sensitive to context in ethical judgments is one of the hardest things to get the program to do (it's also hard for humans), the lower scores are, perhaps, to be expected.

Since the expert assigned a mark to each point of similarity and difference generated by the program, we have been able to evaluate how well specific Comparison Rules functioned. For each of 58 rules, we assigned scores based on the expert's marks (i.e., X = 0, check- = 1, check = 2 check+ = 3). We found that the expert scored highly TRUTH-

TELLER's ability to point out differences based on: unshared compelling reasons not to tell the truth, life critical consequences of an action, varying levels of duty to tell the truth associated with particular roles and relationships among the participants, and untried alternative actions. The latter two are very significant because reasoning about roles, relationships, and alternative actions helps TT make context sensitive comparisons. On the other hand, the expert did not seem satisfied with the program's ability to point out differences based on: unshared reasons to tell the truth or not (the rule fired a lot of times but did not always make a sensible contribution) and the existence of affected others. From the expert's comments, we conclude that he felt the program was not successful in tying the existence of the affected others to a specific argument for or against an action. This also goes to context sensitivity. We believe we can address this point.

The evaluation has been conducted at an early stage of development of the TRUTH-TELLER program. A more formal evaluation would require obtaining other experts' evaluations of the same data. Since we are at a preliminary stage, however, we prefer to see whether this expert's evaluation and comments can lead to improvements in the knowledge representation. For instance, the expert commented generally that the comparison texts lacked an "organizing roadmap" and a recommended final decision "which could guide thinking and in terms of which the similarities, differences, and ethical principles could be marshalled." As per his suggestion, we are reorganizing the comparison text around specific conclusions and experimenting with various techniques for formulating more pointed arguments supporting a conclusion.

In conclusion, the evaluation shows that we have developed an AI CBR knowledge representation which can very nearly simulate intelligent comparisons of practical ethical problems, at least for 23 cases in the field of truth telling. We anticipate a final case base of between 50 and 75 cases. We plan to use this knowledge representation for a cognitive science investigation of the differences between novice and experienced reasoning about ethical dilemmas. As previously noted, we are collecting and analyzing protocols of high school and ethics graduate students' arguments about the same cases as in TRUTH-TELLER's database. We hope to use our knowledge representation to duplicate and explain some of the observed differences in novice and experienced reasoning and to model some of the arguers' specific differences of opinion. We also hope to model arguments that a proposed resolution is inconsistent with an arguer's previous arguments about similar cases. This investigation will lead, we hope, to a realistic model of ethical decision-making or, at least, of reasonable ethical case-based argumentation upon which to base a tutoring system to improve novices' abilities to recognize and analyze practical ethical dilemmas.

References

Ashley, K. D. (1990) *Modeling Legal Argument: Reasoning with Cases and Hypotheticals.* MIT Press, Cambridge,. Based on PhD. Dissertation, University of Massachusetts, 1987, COINS Technical Report No. 88-01.

Bareiss, E. R. (1989) *Exemplar-Based Knowledge Acquisition - A Unified Approach to Concept Representation, Classification, and Learning.* Academic Press, San Diego, CA. Based on PhD thesis, University of Texas, 1988.

Beauchamp, T. and McCullough, L. B. (1984) *Medical Ethics: The Moral Responsibilities of Physicians.* Prentice-Hall, Englewood Cliffs, NJ.

Bok, S. (1989) *Lying.* Random House, Inc. Vintage Books, New York.

Edelson, D. C. (1992) When Should a Cheetah Remind You of a Bat? Reminding in Case-Based Teaching. In *Proceedings AAAI-92*, pages 667-672. American Association for Artificial Intelligence. San Jose, CA.

Jonsen, A. R. & Toulmin, S. (1988) *The Abuse of Casuistry: A History of Moral Reasoning.* University of CA Press, Berkeley.

Kass, A. M. ,Leake, D., and Owens C. C. (1986) Swale: A Program that Explains. In Roger C. Schank, editor, *Explanation Patterns: Understanding Mechanically and Creatively.* Lawrence Erlbaum Associates, Hillsdale, NJ.

MacGregor, R. (1991) The Evolving Technology of Classification-Based Knowledge Representation Systems. In John F. Sowa, editor. *Principles of Semantic Networks: Explanations in the Representation of Knowledge*, pages 385-400. Morgan Kaufmann, San Mateo, CA.

Rissland, E. L., Skalak, D. B. & Friedman, M. T. (1993) BankXX: A Program to Generate Argument through Case-Based Search. In *Fourth International Conference on Artificial Intelligence and Law*, Vrie Universiteit, Amsterdam

Schaffner, K. F. (1990) Case-Based Reasoning in Law and Ethics. Presentation at the "Foundations of Bioethics" Conference. Hastings Center.

Strong, C. (1988) Justification in Ethics. In Baruch A. Brody, editor, *Moral Theory and Moral Judgments in Medical Ethics*, pages 193-211. Kluwer Academic Publishers, Dordrecht.

Veloso, M. M. (1992) *Learning by Analogical Reasoning in General Problem Solving.* PhD thesis, Carnegie Mellon University. Technical Report No. CMU-CS-92-174.

Woods, W. A. and Schmolze, J. G. (1990) The KL-ONE Family. Technical Report TR-20-90, Center for Research in Computing Technology, Harvard University, Cambridge, MA.