Title

Case-Based Comparative Evaluation in TRUTH-TELLER

Permalink

https://escholarship.org/uc/item/7pj939zh

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 17(0)

Authors

McLaren, Bruce M.

Ashley, Kevin D.

Publication Date

1995

Peer reviewed

Case-Based Comparative Evaluation in TRUTH-TELLER

Bruce M. McLaren
University of Pittsburgh
Intelligent Systems Program
Pittsburgh, Pennsylvania 15260
bmm@cgi.com

Kevin D. Ashley
University of Pittsburgh
Intelligent Systems Program
Pittsburgh, Pennsylvania 15260
ashley@vms.cis.pitt.edu

Abstract

Case-based comparative evaluation appears to be an important strategy for addressing problems in weak analytic domains, such as the law and practical ethics. Comparisons to paradigm, hypothetical, or past cases may help a reasoner make decisions about a current dilemma. We are investigating the uses of comparative evaluation in practical ethical reasoning, and whether recent philosophical models of casuistic reasoning in medical ethics may contribute to developing models of comparative evaluation. A good comparative reasoner, we believe, should be able to integrate abstract knowledge of reasons and principles into its analysis and still take a problem's context and details adequately into account. TRUTH-TELLER is a program we have developed that compares pairs of cases presenting ethical dilemmas about whether to tell the truth by marshaling relevant similarities and differences in a context sensitive manner. The program has a variety of methods for reasoning about reasons. These include classifying reasons as principled or altruistic, comparing the strengths of reasons, and qualifying reasons by participants' roles and the criticality of consequences. We describe a knowledge representation and comparative evaluation process for this domain. In an evaluation of the program, five professional ethicists scored the program's output for randomly-selected pairs of cases. The work contributes to context sensitive similarity assessment and to models of argumentation in weak analytic domains.

Introduction

Case-based comparative evaluation appears to be an important strategy for addressing problems in weak analytic domains. Such domains require the construction of arguments or explanations to justify decisions but cannot support the use of deductive methods or formal proofs to derive correct answers. Comparison to paradigm, hypothetical, or past cases can help a reasoner make decisions about a problem situation. For instance, in the legal domain, lawyers form arguments, at least in part, by analogizing to previously adjudicated cases and hypotheticals (Ashley, 1990). Practical ethical reasoning — and, in particular, truth telling — is another weak analytic domain in which such a comparative evaluation model (CEM) could prove useful. A reasoner faced with an ethical dilemma could select paradigmatic, hypothetical, and past cases, compare them to the problem, construct arguments identifying the critical reasons justifying their importance by drawing analogies to the cases, and evaluate the competing arguments to resolve the dilemma.

Medical ethicists have recently revived a case-based (i.e., casuistic) approach to practical ethical reasoning in which problems are compared to past or paradigmatic cases (Strong, 1988, Jonsen and Toulmin, 1988). For instance, one ethicist proposed the following steps when one is faced with a moral problem:

1. Identify middle-level principles and role-specific duties pertinent to the 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 a paradigm case in which the option would be justifiable. Paradigms can be actual or hypothetical cases. Identify the middle-level principle which would provide that justification.
5. Compare the case at hand with paradigm cases. Determine which paradigms it is "closest to" in terms of the presence of morally relevant factors (Strong, 1988).

Computationally realizing a model like the above is interesting because it addresses important goals that case-based reasoning (CBR) has not yet modeled such as: (1) symbolically comparing problems and paradigmatic cases in terms of the underlying principles, reasons, and actions and (2) adequately accounting for a problem's specific contextual circumstances in deciding how to resolve conflicting reasons and principles.

A key problem in decision making with principles, reasons, and cases is this: humans find it hard to integrate different kinds of knowledge which vary from the very abstract (principles) through an intermediate range (reasons) to the very specific (cases). Cognitive psychological evidence has shown, for instance, that gender and developmental differences affect a reasoner's ability to account for contextual features in applying general principles in moral decision making (Gilligan, 1982; Johnston, 1988). Some people are better than others at resolving ethical dilemmas, a process which requires one to take into account the problem's particular factual circumstances, qualify reasons based on their criticality and the participants' roles, relationships and interests, and consider possible alternative actions.

Although AI / CBR programs have illustrated a variety of techniques for modeling comparative evaluation (see, for example, Bareiss, 1989; Golding and Rosenbloom, 1991), to our knowledge, a knowledge representation and inference technique has not been developed which integrates reasons, principles, and cases. Like (Edelson, 1992), we represent principles which apply to cases at various levels of abstraction. In CABARET (Rissland and Skalak, 1991), the circumstances also included the arguer's viewpoint and various argument moves associated with broadening or restricting a statutory predicate. In (Rissland, Skalak, and Friedman, 1993), legal rules, theories, standard stories, and family resemblance have been integrated into case retrieval and argumentation. We, however, intend for our program to use principles and to reason about reasons differently from these programs. In particular, we are attempting to enable a program to use case comparison to decide whether principles and reasons apply more or less strongly in one case than another.

In this paper, we report on our progress toward developing a CEM in the domain of practical ethics. TRUTH-TELLER (TT) is a program we have developed to compare pairs of cases presenting ethical dilemmas about whether to tell the truth. TT's comparisons point out ethically relevant similarities and differences (i.e., reasons for telling or not telling the truth that (1) apply to both cases, (2) apply more strongly in one case than another or (3) apply only to one case). In developing the knowledge representation for symbolic case comparison, we have adhered to an approach of repeated development and formative evaluation (Ashley and McLaren, 1994). Although TT does not implement Strong's CEM, it provides the kind of case comparison that would be essential to compare a problem and a paradigmatic case (Step 5 in Strong's procedure.)

TT's Case Comparison Method

Having accepted as input representations of two cases to be compared, TT's case comparison method proceeds in four sequential phases:

(1) **The Alignment Phase.** Aligning reasons means building a mapping between the reasons in two cases. The initial phase of the program "aligns" the semantic representations of the two input cases by matching similar reasons, actor relations, and actions, by marking reasons that are distinct to one case, and by noting exceptional reasons in one or both of the cases.

(2) **The Qualification Phase.** Qualifying a reason means identifying special relationships among actors, actions, and reasons that augment or diminish the importance of the reasons. Heuristic rules strengthen and weaken individual reasons and actions. Attributes such as altruism, selfishness, and high criticality are applied as qualifiers to reasons and actions. Also, the alignment links between reasons, relations, and actions of the two opposing cases are tagged with qualifying information based on the participants' roles, reason types, and untried alternatives.

(3) **The Marshaling Phase.** Marshaling reasons means selecting particular reason similarities and differences to emphasize in presenting an argument that (1) one case is stronger than the other with respect to a conclusion, (2) the cases are only weakly comparable, or (3) the cases are not comparable at all. Marshaling serves *rhetorical* criteria for deciding how to integrate facts, reasons, and justifications into a convincing output.

(4) **The Interpretation Phase.** The final phase of the program generates the comparison text by interpreting the activities of the first three phases. The purpose of this phase is to generate prose that a nontechnical human evaluator can understand.

The program employs various knowledge structures to support its algorithm including semantic networks that represent the truth telling episodes, a relations hierarchy, and a reasons hierarchy. All structures are implemented in LOOM (MacGregor, 1990).

Each truth telling episode includes representations for the actors (i.e., the truth teller, truth receiver, and others affected by the decision), relationships between the actors (e.g., familial, professional, seller-customer), the truth teller's possible actions (i.e., telling the truth, not telling the truth, or taking some alternative action) and reasons that support the possible actions.

The relations hierarchy is a taxonomy of approximately 80 possible relationships among the actors in a truth telling episode. Relationships include familial, commercial, and acquaintance relations. More abstract relationship types include high-trust, minimal-trust, and authority relations.

The reasons hierarchy represents possible rationales for taking action. Based on the formulation in (Bok, 1989), the hierarchy employs, at its top tier, four general reasons for telling the truth or not: fairness, veracity, producing of benefit, and avoiding harm. All other reasons are descendants of one of these top-level reason types. Each reason also has three other facets, criticality, if altruistic, and if principled, each of which is important to ethical decision-making.

An Example of TT's Case Comparison Method

We now illustrate TT's case comparison method by tracing an example output of the program. We guide the reader through the four phases, focusing on the underlined portion of the comparison text of Figure 1.

TRUTH-TELLER is comparing the following cases:

CASE 1: Victor is a young lawyer running his own business. A client requires a complex legal transaction that Victor has never done before. Should Victor tell the client about his inexperience in this matter?

CASE 2: Terry coaches a little league team. Half way through the season Terry discovers that the star player, Sammie, is three months over age. Should Terry ignore this information?

TRUTH-TELLER's analysis:

Victor and Terry are faced with similar dilemmas. They abstractly share reasons to both tell the truth and not tell the truth. The episodes abstractly share one reason to tell the truth. Victor and Terry share the general reason to provide fairness. Victor has the reason to provide sales information so that a consumer can make an informed decision for Victor's client and to disclose professional inexperience for Victor's client, while Terry has the reason to enforce rules of a game that have been violated by Sammie for the players on other teams.

The two cases also abstractly share a reason to not tell the truth. Victor and Terry share the general reason to produce benefit. Victor has the reason to enhance professional status and opportunities and to realize a financial gain for himself, while Terry has the reason to attain a competitive advantage for Sammie's teammates, Sammie and himself.

However, these quandaries are distinguishable. Arguments can be made for both Victor and Terry having a stronger basis for telling the truth.

On the one hand, one could argue that telling the truth is better supported in Victor's case. The reason 'to provide fairness', a shared reason for telling the truth, is stronger in Victor's case, because it involves a higher level of trust and duty between Victor and Victor's client.

On the other hand, one could also argue that Terry has a more compelling case to tell the truth. First, Terry may tell the truth to provide an example of honesty for children for the players on other teams, Sammie's teammates and Sammie. Additionally, the shared reason for not telling the truth 'to produce benefit' is weaker in Terry's case, because the competitive advantage to Terry and the team is probably small. Third, Terry's motivations for telling the truth appear to be fully principled (i.e. 'One should make a special effort to be truthful when a child may be influenced towards honesty.', 'The rules of a game should be followed.'). However, Victor's motivations are not all principled (e.g. Victor has the unprincipled reason 'to establish goodwill for future benefit'). Finally, Terry appears to have purely altruistic reasons for telling the truth, thus strengthening his case to tell the truth. On the other hand, Victor has selfish motivation to tell the truth. Victor may tell the truth to establish goodwill for future benefit.

Figure 1: TT's Output Comparing Victor's and Terry's Cases

The program starts by accepting semantic representations of each of the cases. Figure 2 depicts the semantic representations of the Victor and Terry cases. In Terry's case Terry is the truth teller, since it is he who is confronted with the decision to tell the truth or not. The league authority will hear the truth, should Terry decide to divulge it, and thus is the truth receiver. Finally, Sammie, Sammie's teammates, and the players on other teams are affected others, since they would be affected in some way by disclosure.

The semantic representation also contains a set of possible actions that the truth teller could take and reasons supporting or justifying 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 informa-

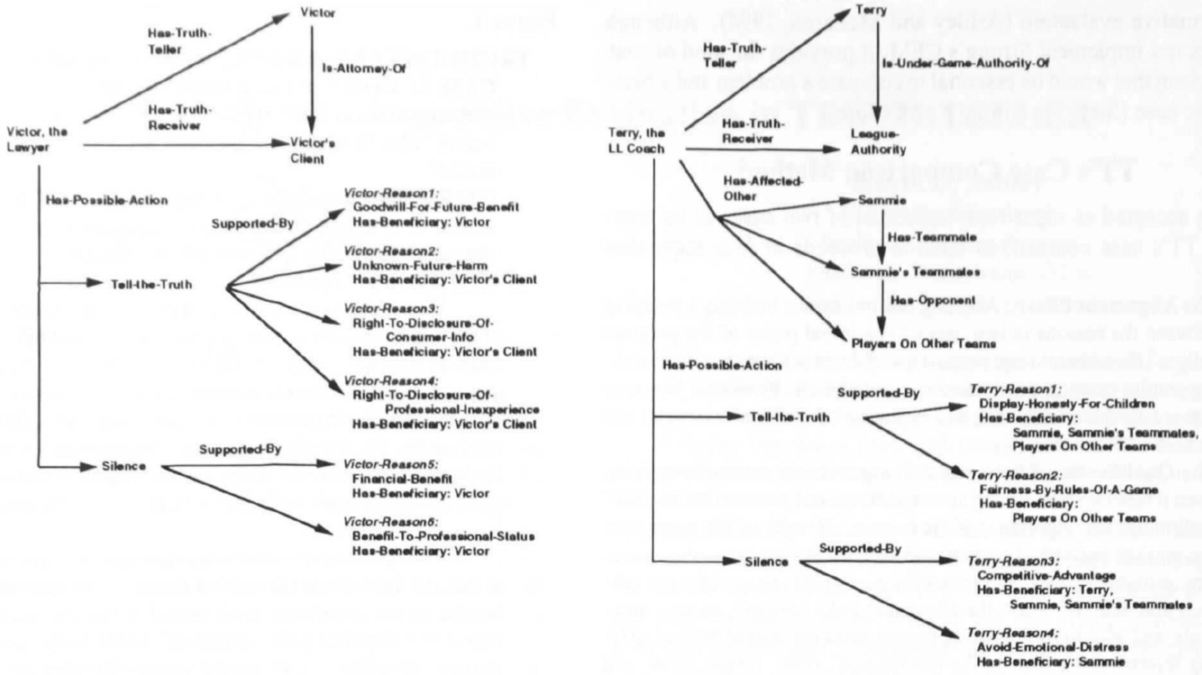
Victor
Has-Truth-Teller
Is-Attorney-Of
Has-Truth-Receiver
Victor, the Lawyer
Victor's Client
Has-Possible-Action
Supported-By
Victor-Reason1: Goodwill-For-Future-Benefit Has-Beneficiary: Victor
Victor-Reason2: Unknown-Future-Harm Has-Beneficiary: Victor's Client
Tell-the-Truth
Victor-Reason3: Right-To-Disclosure-Of-Consumer-Info Has-Beneficiary: Victor's Client
Victor-Reason4: Right-To-Disclosure-Of-Professional-Inexperience Has-Beneficiary: Victor's Client
Supported-By
Silence
Victor-Reason5: Financial-Benefit Has-Beneficiary: Victor
Victor-Reason6: Benefit-To-Professional-Status Has-Beneficiary: Victor
Terry
Has-Truth-Teller
Is-Under-Game-Authority-Of
Has-Truth-Receiver
Terry, the LL Coach
League-Authority
Has-Affected-Other
Sammie
Has-Teammate
Sammie's Teammates
Has-Opponent
Players On Other Teams
Has-Possible-Action
Supported-By
Tell-the-Truth
Terry-Reason1: Display-Honesty-For-Children Has-Beneficiary: Sammie, Sammie's Teammates, Players On Other Teams
Terry-Reason2: Fairness-By-Rules-Of-A-Game Has-Beneficiary: Players On Other Teams
Silence
Supported-By
Terry-Reason3: Competitive-Advantage Has-Beneficiary: Terry, Sammie, Sammie's Teammates
Terry-Reason4: Avoid-Emotional-Distress Has-Beneficiary: Sammie

Figure 2: Representation of Victor's Case (left) and Terry's Case (right)

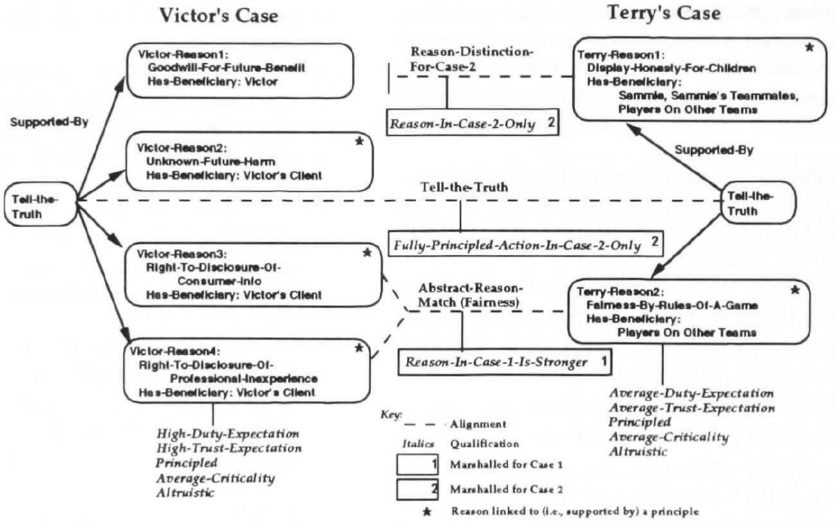
Victor's Case
Terry's Case
Victor-Reason1: Goodwill-For-Future-Benefit Has-Beneficiary: Victor
Supported-By
Victor-Reason2: Unknown-Future-Harm Has-Beneficiary: Victor's Client
Tell-the-Truth
Victor-Reason3: Right-To-Disclosure-Of-Consumer-Info Has-Beneficiary: Victor's Client
Victor-Reason4: Right-To-Disclosure-Of-Professional-Inexperience Has-Beneficiary: Victor's Client
High-Duty-Expectation
High-Trust-Expectation
Principled
Average-Criticality
Altruistic
Reason-Distinction-For-Case-2
Reason-In-Case-2-Only 2
Tell-the-Truth
Fully-Principled-Action-In-Case-2-Only 2
Abstract-Reason-Match (Fairness)
Reason-In-Case-1-Is-Stronger 1
Terry-Reason1: Display-Honesty-For-Children Has-Beneficiary: Sammie, Sammie's Teammates, Players On Other Teams
Supported-By
Tell-the-Truth
Terry-Reason2: Fairness-By-Rules-Of-A-Game Has-Beneficiary: Players On Other Teams
Average-Duty-Expectation
Average-Trust-Expectation
Principled
Average-Criticality
Altruistic
Key:
— — Alignment
Italics Qualification
1 Marshalled for Case 1
2 Marshalled for Case 2
★ Reason linked to (i.e., supported by) a principle

Figure 3: A Portion of the Victor and Terry Comparison after Alignment, Qualification, and Marshaling

tion). In both Victor's and Terry's case, the choice is between telling the truth and remaining silent. Other episodes in TT's case base involve outright lying. There is also the possibility that an alternative action could be taken, although not so in Victor's or Terry's case. For instance, in some cases the truth teller may have the option to approach an affected other before either telling the truth or lying to the truth receiver[1]

In our knowledge representation actions are supported by reasons; a reason is treated as a rationale for taking an action. For example, a rationale for Victor's telling the truth is to protect his client's right of disclosure. On the other hand, a rationale for Victor's silence is the possibility of financial gain. Notice that rationales do not need to be selfless or principled; in fact, one of the goals of this work is to imbue the program with the capability of distinguishing between selfless and selfish reasons and between principled and unprincipled reasons for action.

The program first performs the Alignment phase. The dashed lines in Figure 3 depict alignments between the Victor and Terry representations. At the top of the diagram, Terry's reason 'display honesty for children' is determined to have no counterpart in the Victor representation (i.e., it is a clear distinction); thus it is "misaligned" and labelled as a reason distinction. Moving downward, the 'tell the truth' actions of each case are aligned with one another, since they represent the same action. Finally, at the bottom of the diagram, Victor and Terry have reasons that abstractly match and are thus aligned with one another (i.e., Victor has two right of disclosure reasons, while Terry has the reason to uphold the rules of a game). These reasons match in the reason hierarchy at the level of 'fairness.'

The program next commences the Qualification phase. The italicized text in Figure 3 represents the qualifications that are applied to the comparison. The first step is to qualify the individual objects to

[1] Consider the following case. "Rick's father is having an affair. Rick's mother is unaware of it. Should Rick tell his mother?" Before Rick "blows the whistle" on his father, he could discuss the issue with him.

reflect the strength of reasons (or actions) relative to a case itself, irrespective of its comparison to the other case. For instance, the reason 'right to disclosure of professional inexperience' is tagged (1) as having high duty and trust expectations, since it involves a relationship between a professional advisor and a client, (2) as being principled, because an ethical principle supports the reason (Notice the '*' in the upper right region of the reason. This connotes a principle link. The particular principle in this case is 'In a situation in which a professional is being depended upon, information regarding the inexperience of that professional should be disclosed.'), (3) as having average criticality, since no comment was made in the story about critical consequences, and (4) as being altruistic, since the reason is to the benefit of Victor's client and not to Victor himself. The corresponding reason in Terry's case (i.e., 'fairness by rules of a game') is tagged likewise, with the exception that its trust and duty levels are labelled as average, since the relationship between Terry and the players on the other teams is not one that would typically involve a high level of trust or duty.

The second step of Qualification is to qualify alignments to reflect relative differences between reasons (or actions) across the cases. In Figure 3 there are three alignment qualifications. The first one is the reason distinction misalignment which is tagged as being a reason found only in case 2. Second, the alignment between Victor's 'tell the truth' action and Terry's 'tell the truth' action is tagged as being fully principled, and thus stronger, on Terry's side, since both of Terry's reasons for telling the truth have associated ethical principles, while one of Victor's rationales, the 'goodwill for future benefit' reason, is unprincipled (Notice there is no "*" associated with this reason in Figure 3). Finally, the abstract reason match at the bottom of Figure 3 is tagged as stronger for Victor. This is so because of the high trust and duty involved with Victor's 'right to disclosure' reasons as opposed to the average trust and duty involved with Terry's 'fairness by rules of a game' reason[2].

Next, the program begins the Marshaling phase. Its first marshaling task is to assign the case comparison to one of five possible comparison contexts. The five comparison contexts are defined as follows:

1. *Comparable-Dilemmas/Reason-Similarity*, if the cases present similar dilemmas. i.e., the reasons supporting both telling the truth and not telling the truth are similar either in an identical or abstract way,
2. *Comparable-Dilemmas/Criticality-Similarity*, if the cases are similar due to the critical nature of possible consequences,
3. *Comparable-Reasons*, if the cases share a similar reason or reasons for either telling the truth or not telling the truth but not for both possible actions,
4. *Incomparable-Dilemmas/Reason-Difference*, if the cases do not have any reasons supporting like actions that are similar, and
5. *Incomparable-Dilemmas/Criticality-Difference*, if the cases are incomparable due to a difference in the criticality of the possible consequences.

The Victor/Terry comparison is classified as an instance of the *Comparable-Dilemmas/Reason-Similarity* comparison context, since it has abstract reasons to both tell the truth and not tell the truth. After classifying the comparison, the program marshals information that is appropriate to the classified context. There are two general categories of information that are marshaled, the *comparison focus* (i.e., information that is to be the initial focus of comparison and is typically the most important information to draw attention to) and the *distinguishing information* (i.e., information that contrasts to the comparison focus). For instance, for the *Comparable-Dilemmas/Reason-Similarity* comparison context the program marshals the similar reasons and relations as the comparison focus and then, to distinguish the cases, marshals the information that supports arguing the relative merits of telling the truth in the two cases.

Now let us return to Figure 3 to explain how the data in the figure is marshaled. Marshaled information is enclosed in a box with a number 1 or 2, corresponding to whether the information was marshaled in support of an argument for case 1 (Victor's case) or case 2 (Terry's case). Only the marshaled distinguishing information is depicted in the diagram. It is interesting to note, however, that the abstract reason match in Figure 3 is actually marshaled as part of both the comparison focus and the distinguishing information. This is so because the reason match is a similarity, but qualification has also revealed it as giving rise to a distinction at a more detailed level. Note that the qualifier *Reason-In-Case-1-Is-Stronger* strengthens the case for Victor telling the truth, relative to Terry. This marshaled data corresponds to the underlined text in paragraph four of Figure 1 (the sentence beginning "The reason 'to provide fairness'..."), the paragraph that argues Victor's truthtelling strengths.

The remaining marshaled information in Figure 3 supports the argument that Terry has a more compelling case to tell the truth. Terry's reason to 'display honesty for children' is marshaled as a strength of Terry's case relative to Victor's, because it is a misaligned reason distinction (i.e., it provides a justification for Terry to tell the truth that is unshared by Victor). This corresponds to the underlined sentence in Figure 1 beginning "First, ..." Finally, the program marshals the qualifier, *Fully-Principled-Action-In-Case-2-Only*, from the aligned actions in Figure 3. This also supports Terry's case relative to Victor's, since it shows a strength for telling the truth that exists for Terry but not Victor. This final marshaled information corresponds to the text beginning "Third, ..."

The final phase of the program, Interpretation, generates the natural language depicted in Figure 1 by traversing subgraphs of an augmented transition network (ATN) (McKeown, 1985). Each comparison context is represented by a different subgraph in the ATN. As the program traverses the ATN it generates rhetorical predicates, the basic units of discourse. Each rhetorical predicate essentially maps to a sentence in the comparison text.

TT's sensitivity to comparison context is illustrated by contrasting the output of Figure 1 with the output of Figure 4. The comparison text in Figure 4 is an example of the context *Incomparable-Dilemmas/Criticality-Difference*. In this context, the program first focuses attention on the critical difference between the cases (i.e., the possible consequences), rather than on the similarities, as was done in the comparison of Figure 1. Also, there is no reason to argue the relative merits of telling the truth in this particular comparison, as was done in Figure 1, since the cases are not comparable. Instead, the program contrasts the two cases by reciting the less critical reasons associated with Victor's case.

TRUTH-TELLER is comparing the following cases:

CASE 1: Victor is a young lawyer running his own business. A client requires a complex legal transaction that Victor has never done before. Should Victor tell the client about his inexperience in this matter?

CASE 2: Josh encounters a man whom he recognizes from the newspaper as a murderer-at-large. The paper stated that the murderer had killed two of his three ex-wives. Josh's neighbor, Judy, is the final ex-wife of the murderer. The man asks Josh where Judy lives.

[2] A Reason X is said to be stronger than a Reason Y iff Reason X has a higher criticality than Reason Y OR Reason X has at least one qualifier (i.e. trust, duty, altruism, principled) that is stronger than Reason Y's AND Reason X has no qualifier that is weaker than Reason Y's.

Should Josh tell the murderer where Judy lives?

TRUTH-TELLER's analysis:

The possible consequences of the quandaries faced by Victor and Josh are qualitatively different. The consequences of Josh's decision are life critical. Thus, the dilemmas are difficult to compare. Josh has the reason to not tell the truth because Judy's life is in peril. In addition, Josh has the reason to not tell the truth because Judy's family would be quite distressed if the murder occurs.

Victor's reasons for both telling and not telling the truth are far less critical than Josh's reasons. On the one hand, Victor may tell the truth to provide sales information so that a consumer can make an informed decision. In addition, Victor may tell the truth to disclose professional inexperience for Victors client. Third, Victor may tell the truth to establish goodwill for future benefit for himself. Finally, Victor has the reason to tell the truth to avoid an unknown future harm for Victors client. On the other hand, Victor has the reason to not tell the truth to enhance professional status and opportunities for himself. Additionally, Victor may not tell the truth to realize a financial gain for himself.

Figure 4: TT's Output Comparing Victor's and Josh's Cases

To summarize, the extended example shows how TRUTH-TELLER generates context sensitive case comparisons using alignment, qualification, and marshaling. It "reasons about reasons" by aligning cases according to similarities and differences, and qualifying cases in various ways including tagging reasons as altruistic, principled, critical, high trust, high duty, etc. and tagging alignments with relative strengths. The comparison context dictates the strategy the program employs to marshal the relevant similarities and differences. Finally, the interpretation phase uses the marshaled information to generate a comparison text.

The Evaluation

Our goal was to obtain some assurance that TT generated case comparisons that expert ethicists would regard as appropriate. Our experimental design for this formative evaluation was to poll the opinions of five expert ethicists as to the reasonableness, completeness, and context sensitivity of a relatively large sampling of TT's case comparisons.

We divided the evaluation into two parts. The first experiment presented the experts with twenty comparison texts TT generated for pairs of cases randomly selected. We also added two comparison texts generated by humans (a medical ethics graduate student and a law school professor). The evaluators were informed that some texts were generated by humans and some were generated by a computer program; however, they were not told specifically which or how many comparisons were by humans and which were done by the program. The second experiment presented the experts with five comparison texts in which TT compared the same case to five different cases. For each experiment, the evaluators were instructed: "In performing the grading, we would like you to evaluate the comparisons as you would evaluate short answers written by college undergraduates. ... Please focus on the substance of the comparisons and ignore grammatical mistakes, awkward constructions, or poor word choices (unless, of course, they have a substantial negative effect on substance.)" We also instructed the experts to critique each of the comparison texts.

In the first experiment, we instructed the experts to assign three grades to each of the twenty-two comparison texts, a separate grade for reasonableness, completeness, and context sensitivity. The scale for each grading dimension was 1 to 10, to be interpreted by the evaluators as follows: for reasonableness, 10 = very reasonable, sophisticated; 1 = totally unreasonable, wrong-headed; for completeness, 10 = comprehensive and deep; 1 = totally inadequate and shallow; for context sensitivity, 10 = very sensitive to context, perceptive; 1 = very insensitive to context.

The results of the first experiment were as follows. The mean scores across the five experts for the twenty TT comparisons were R = 6.3, C = 6.2, and CS = 6.1. Figure 5 shows the maximum, minimum and mean scores per comparison for all three of the dimensions. By way of comparison, the mean scores of the two human-generated comparisons were R = 8.2, C = 7.7 and CS = 7.8. Not surprisingly, one of the human comparisons, number 16, attained the highest mean on all three dimensions (R = 9; C = 8.8; CS = 8.8). Two of the program generated comparisons (numbers 2 and 14), however, were graded higher on all three dimensions than the remaining human comparison (number 22).

In the second part of the evaluation, we wanted to focus on the program's sensitivity to context. To achieve this, we asked the experts to grade five additional TRUTH-TELLER comparisons. These comparisons all involved one case -- the Victor case -- repeatedly compared to a different second case (i.e., a one-to-many comparison). The TT comparisons discussed in this paper (i.e., Figures 1 and 4) were included in the second part of the experiment. For this part of the evaluation, the experts were asked to grade all five comparisons as a set, assigning three scores, one for each of the three dimensions (i.e., reasonableness, completeness, and context sensitivity).

The results of the second part of the evaluation were as follows. The mean across the five experts was R = 6.7; C = 6.9; CS = 7.0. Notice that the program fared better on the context sensitivity dimension than on the other two dimensions. This contrasts to the first part of the experiment in which the mean CS score was the lowest of the three dimensions. Also, notice that the scores of all three dimensions were improved slightly over the first experiment.

Discussion and Conclusions

Our results should be viewed in light of our goals and the experimental design in this formative evaluation. We solicited expert opinions about the adequacy of TRUTH-TELLER's comparison texts in order to assess whether our knowledge representation and reasoning techniques were appropriate to the domain task and to obtain critiques identifying areas for improvement. Our primary intention was to determine if TT's comparisons were at least "within range" of that of humans and to determine the ways in which our model could be improved. We interpret the results as indicating that TRUTH-TELLER is somewhat successful at comparing truth telling dilemmas. We included the two human-generated texts as a calibration of the experts' scores; we are encouraged that some of the program's grades were higher than those assigned to texts written by post graduate humans.

On the other hand, our experiment does not involve an adequately sized sampling of human comparisons nor did we present the experts with outputs in which TRUTH-TELLER and humans generated comparison texts for the same pairs of cases. Quite simply, we felt it was premature to adopt this kind of experimental design for a formative evaluation.

The second part of the experiment attempts to address whether TRUTH-TELLER is competent at marshaling comparisons in a context sensitive manner. We believe that the higher scores in the second part of the experiment are due, at least in part, to a keener appreciation of the program's task; that is, it was easier for the evaluators

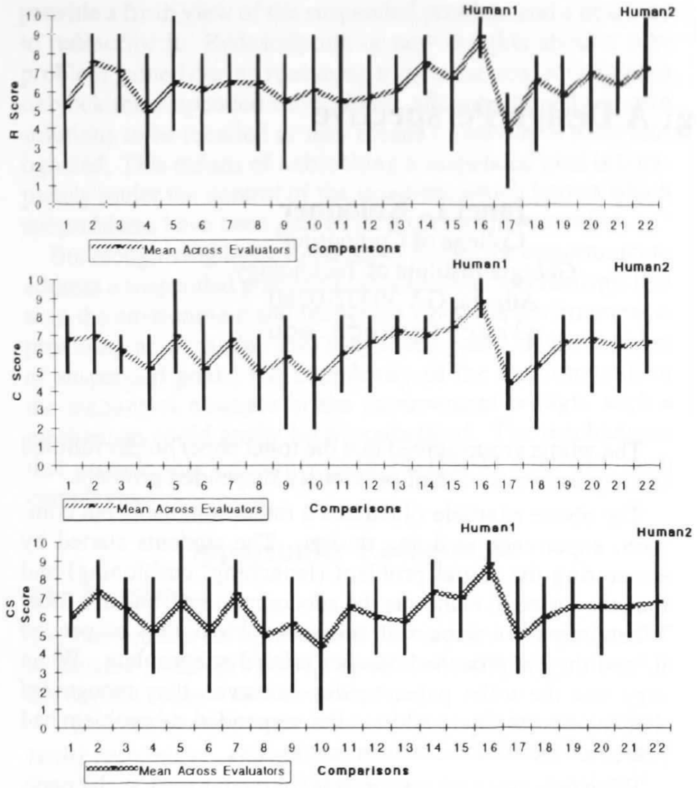

Figure 5: Max, Min, and Mean Values for R (top), C (middle), and CS (bottom) Scores in Experiment #1

to recognize TT's sensitivity to context in the one-to-many experiment. We have shown a flavor of this by providing and discussing two of the five comparisons (i.e., Figures 1 and 4) from the one-to-many the experiment. Figures 1 and 4 illustrate two points on the range of comparisons that the program can draw upon in comparing the same case to other cases.

We also invited the evaluators to critique the texts. The evaluators were in general agreement about some points (i.e., some comments appeared multiple times, across multiple evaluators). For instance, several evaluators questioned TRUTH-TELLER's lack of hypothetical analysis (i.e., the program makes immutable assumptions, eschewing "what if" analysis.). Addressing this would require a program imbued with a more elaborate representation of reasons, actions, and actors. We are considering modifying cases to consider "factors" in a heuristic manner, similar to that done in (Ashley, 1990). The evaluators also had the tendency to question abstract reason matches. For instance, one evaluator questioned the program making an abstract connection 'producing benefit' between the reasons 'to enhance professional status' and 'to realize a financial gain.' This points, perhaps, to disagreement about the structure of the reason hierarchy (e.g., Is there a more specific connection between these reasons, such as 'Gaining selfish job benefits'?). It may also be that, when reflecting on ethical dilemmas, humans typically think in terms of exact matches of reasons or principles. We need to determine under what circumstances, if any, an abstract match is interesting or important.

It is clear that our representation and comparative analysis could benefit from consideration of the evaluators' critiques, such as those above, and we plan to carefully review these before developing the next iteration of the program.

In conclusion, the evaluation confirmed that TRUTH-TELLER makes mostly reasonable comparisons (although not as good as humans), can make comparisons over a range of cases, and is sensitive to comparison context. Further, we believe that the evaluation has shown that TT's AI / CBR knowledge representation and marshaling process provides a solid foundation for the development of a comparative evaluation model in the truth telling domain. We intend to work on developing a CEM like those proposed by Strong and Jonsen/Toulmin.

Acknowledgements

This work is supported by the Andrew W. Mellon Foundation. We are grateful to Dr. Reidar K. Lie, Dr. James L. Nelson, Dr. Carson Strong, Dr. Matt Keefer, and Mr. Ross Parham for participating in our formative evaluation.

References

Ashley, K. D. (1990). *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. MIT Press, Cambridge. Based on doctoral dissertation, University of Massachusetts, 1987.

Ashley, K. D. and McLaren, B. M. (1994). A CBR Knowledge Representation for Practical Ethics. In the *Second European Workshop on Case-Based Reasoning*. Chantilly, France. To be Published in M. Keane, editor. *Lecture Notes in Artificial Intelligence*, Springer Verlag: Berlin.

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

Bok, S. (1989). *Lying: Moral Choice in Public and Private Life*. 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 the *Proceedings of AAAI-92*, (pp. 667-672). San Jose, CA.

Gilligan, C. (1982). *In A Different Voice*. Harvard University Press.

Golding, A. R. and Rosenbloom, P. S. (1991). Improving Rule-Based Systems through Case-Based Reasoning. In the *Proceedings of AAAI-91*.

Johnston, D. K. (1988). Adolescents' Solutions to Dilemmas in Fables: Two Moral Orientations —Two Problem Solving Strategies. In Gilligan C. et al (Eds.), *Mapping the Moral Domain* (pp. 49-72). Cambridge, MA: Harvard University Press.

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

MacGregor, R. (1990) The Evolving Technology of Classification-Based Knowledge Representation Systems. In John F. Sowa, editor. *Principles of Semantic Networks: Explorations in the Representation of Knowledge*. Chapter 13. Morgan Kaufmann Publishers, Inc., San Mateo, CA.

McKeown, K. R. (1985). Discourse Strategies for Generating Natural-Language Text. In *Artificial Intelligence* 27, 1-41. Elsevier Science Publishers B. V. (North-Holland).

Rissland, E. L. and Skalak, D. B. (1991). CABARET: Rule Interpretation in a Hybrid Architecture. In the *Journal of Man-Machine Studies*. 34, 839-887.

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

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