# Context Sensitive Case Comparisons in Practical Ethics: Reasoning about Reasons

Bruce M. McLaren and Kevin D. Ashley
University of Pittsburgh
Intelligent Systems Program
Pittsburgh, Pennsylvania 15260
bmm@cgi.com, ashley@vms.cis.pitt.edu

## Abstract

*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 ethics may contribute to developing models of comparative evaluation. We are also interested in exploring how our work contributes to AI and Law. 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 marshaling process for this domain. In an evaluation of the program, five professional ethicists scored the program's output for randomly-selected comparisons. The work contributes to context sensitive similarity assessment and to models of argumentation in weak analytic domains.*

## 1. Introduction

Comparative evaluation appears to be an important strategy for addressing problems in weak analytic domains. Such domains require the construction of arguments or explanations but cannot use formal proofs to provide correct answers. Comparison to paradigm, hypothetical, or past cases can help a reasoner make decisions about a current 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.

Not all philosophical models of truth telling have involved case comparison. St. Augustine believed that lying could only be pardoned (but not justified) in situations in which lies harm no one and yet save someone from physical harm (Augustine, edited 1952). Sidgwick supported a utilitarian view by suggesting that consequences should be weighed and lying should be judged based on the overall balance between good and evil (Sidgwick, 1907). Unfortunately, these and other approaches have proven largely unsatisfactory in providing guidance for resolving realistic dilemmas. Utilitarian approaches have failed because of the difficulty of assigning and calculating the weights on principles (Beauchamp and McCullough, 1984). Deductive reasoning does not work, because ethical principles are often inconsistent and the conditions under which they apply are not well defined. Situation ethics, or deciding "separately in each particular situation what is the right or obligatory thing to do" (Frankena, 1973 p. 16) does not provide enough guidance on how to make a decision.

Medical ethicists have recently revived another approach to practical ethical reasoning, casuistry, in which problems are compared to past or paradigmatic cases (Strong, 1988, Jonsen and Toulmin, 1988). These comparative evaluation



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. Vincent Aleven, a graduate student in the Intelligent Systems Program, has given us good advice concerning the program. Finally, Gabriele McLaren prepared the final layout.

models integrate decision-making with principles, reasons and cases: comparatively evaluating a problem and paradigmatic, hypothetical and past cases is intended to inform decisions about which principles and reasons apply more strongly in the problem. For instance, one medical 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).

In this paper, we report on some 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 which apply to both cases, and reasons which apply more strongly in one case than another or which apply only to one case). TT is not a full-fledged case-based reasoning (CBR) system. Currently, it does not retrieve cases nor employ any of the cases in its database as paradigm cases. We have built it to develop and test a knowledge representation capable of the kind of case comparisons that would be essential to compare a problem and a paradigmatic case (Step 5 in Strong's procedure) and to ascertain the kinds of relevance criteria which could inform Steps 3, 4 and 5. The knowledge representation for this practical ethical domain includes representations for actions, reasons, principles, truth telling episodes, reason qualifiers, and comparison contexts. In developing the representation for symbolic case comparison, we have adhered to an approach of repeated development and formative evaluation (Ashley and McLaren, 1994).

## 2. The Relationship between Ethical and Legal Reasoning

One of our goals is to explore whether casuistic models of reasoning with ethical cases and principles have anything to offer AI and Law. There are essential similarities between legal and ethical reasoning: "In both common law and common morality, problems and arguments refer directly to particular concrete cases; and they share a common casuistical ancestry, dating from a time when there was not yet any sharp line between law and morality. In both fields, too, the method of finding analogies that enables us to compare problematic new cases and circumstances with earlier exemplary ones, involves the appeal to historical precursors or precedents that throw light on difficult new cases as they occur." (Jonsen and Toulmin, 1988, p. 316). Also, general principles are important in both domains, and there is a dialectical tension between principles and cases in each domain, although perhaps not the same tension. In ethics, the best-matching paradigmatic cases are said to enable a reasoner to resolve conflicting principles (Strong, 1988). In law, the general principles are said to enable determination of which similarities and differences are most important to the selection of the best-matching case (Dworkin, 1977).

Other subtle differences between the domains involve the extent to which a case comparison method is expressly recognized and relied upon and the availability of standard cases. In legal reasoning, the case comparison method is more fully developed and extensively relied upon, at least in Common Law jurisdictions like the United States and the United Kingdom. In these legal systems, there are vast libraries of cases indexed for use in constructing formal justifications such as legal opinions and briefs. Ethicists appear to rely more on memory and written scholarly discussions of cases, and, significantly, few ethical cases are authoritatively determined and most have many more than two possible outcomes[1].

More concretely, our work in developing a CEM in practical ethics is related to several AI and Law programs in so far as it also addresses and extends methods for achieving context sensitivity. In general, context sensitivity in case comparison means knowing what similarities and differences are the most salient in different circumstances: what should a reasoner focus upon and what should it ignore. Different AI and Law programs have expanded the circumstances which a program can take into account in making a determination of salience. In HYPO (Ashley, 1990) and CATO (Aleven and Ashley, 1994), the circumstances included the side argued for, the set of cases being compared and the particular argument move involved (e.g., analogizing, distinguishing, citing a case in various ways as a counter example, various dialectical techniques for emphasizing a strength or downplaying a weakness.) 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. BankXX (Rissland, Skalak, and Friedman, 1993) added legal theories, standard stories and "family resemblance" into the mix. GREBE (Branting, 1991) accounted for judicial determinations in a case about which facts were criterial. Suggestions for incorporating the purposes of legal rules may be found in (Berman and Hafner, 1993). Currently, in its marshaling, TRUTH-TELLER selects salient similarities and differences in light of an overall assessment of two cases' similarity; the program marshals the comparisons differently depending on how close the cases are to one another in terms of categories that ethicists seem to regard as important. TT's heuristics for reasoning about reasons identify other criteria for regarding some similarities and

[1] Many medical ethical cases are drawn, however, from judicial determinations of disputes among hospitals, patients and their families for injunctive relief in connection with medical treatment.

differences as more important than others in terms of underlying principles, criticality of consequences, participants' roles and untried alternatives (Ashley and McLaren, 1994). Ultimately, we hope to use these heuristics to enable TRUTH-TELLER to address another kind of context sensitivity: people making ethical decisions sometimes apply abstract moral principles without adequately taking the particular circumstances of the problem into account. An ethical reasoner needs to be sensitive to the participants' conditions and interests and to the possibility of alternative, less judgemental, solutions to the dilemma.

## 3. TRUTH-TELLER's Case Comparison Method

Having accepted as input representations of two cases to be compared, TRUTH-TELLER'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. The alignment phase is a restricted form of structure mapping (Gentner, 1983, Branting, 1991), in that structural congruence is employed to determine the degree of similarity of the two cases. However, only a few ways of mapping relations and nodes are considered and all nodes and relations are labelled.

(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. During the qualification phase, heuristic production rules qualify or "tag" objects and the alignments between objects in a variety of ways. First, the rules strengthen and weaken individual reasons and actions. Attributes such as altruism, selfishness, and high criticality are applied as qualifiers to reasons and actions. Secondly, 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. These knowledge structures and, in fact, all structures used by the program are implemented using 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. Mid-relationships include familial, commercial, and acquaintance relations. More abstract relationship types include high-trust, minimal-trust, and authority relations. The relations hierarchy is used to infer which relationships are "similar" for purposes of identifying levels of trust and duty that exist between the participants.

Finally, 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. Note that not all reasons are principled; principles are linked to the reasons in the hierarchy that are ethically justified. For instance, the reason 'Avoid-Physical-Harm' has the associated principle 'one should protect oneself and others from serious harm.' On the other hand, the reason, 'Gain-Financial-Benefit' is not supported by any ethical principle. The program accepts an unprincipled reason as a rationale for taking action, but it favors reasons that are principled in a direct comparison.

## 4. An Example of TRUTH-TELLER's Case Comparison Method

We now illustrate TT's case comparison method by following an example of the program in action. Figure 1 presents a sample unedited comparison generated by TT. We first describe the comparison text in its entirety, by describing in general terms what the program does, and then we focus on the underlined portion of the comparison text (the last paragraph of Figure 1) and guide the reader through the four stages of the program.

---

**TRUTH-TELLER is comparing the following cases:**

**CASE 1:** Should Stephanie, a psychology researcher, lie to human subjects about the intent of an experiment in

order to study some aspect of the subject's behavior?

**CASE 2:** Bruce sells radios for a living. His favorite brother, Mark, picks out an expensive model with a history of maintenance problems. Selling this model would mean a big commission to Bruce but a big problem for Mark. Bruce has been doing very well lately, so the commission on this particular radio will not make much difference to his overall financial situation. Should Bruce warn his brother about the potential problems of this radio?

**TRUTH-TELLER's analysis:**

Stephanie and Bruce are faced with similar dilemmas. They abstractly share reasons to both tell the truth and not tell the truth. The cases also share similar relationship contexts. The relationship between Stephanie and the experiment subjects and between Bruce and Mark both involve a high level of duty.

Stephanie and Bruce abstractly share one reason to tell the truth. Both actors share the general reason to protect a right. More specifically, Stephanie has the reason to not trick someone into a disclosure for the experiment subjects, while Bruce has the reason to provide sales information so that a consumer can make an informed decision for Mark.

The two cases also abstractly share a reason to not tell the truth. Stephanie and Bruce share the general reason to produce benefit. Stephanie has the reason to enhance professional status and opportunities for herself, while Bruce has the reason to realize a financial gain for himself.

However, these quandaries also have relevant differences. Arguments can be made for both Stephanie and Bruce having a stronger basis for telling the truth.

On the one hand, there is an argument that telling the truth is better supported in Stephanie's case. First, Stephanie has to decide whether to tell a blatant lie, while Bruce must simply decide whether to remain silent. This fact would tend to put more pressure on Stephanie to tell the truth. Second, Stephanie could possibly acquire information for her research by devising a different experimental procedure. However, according to the story, this action was not taken. Thus, there is a greater onus on Stephanie to be honest.

On the other hand, one could also argue that Bruce has a more compelling case to tell the truth. First, the shared reason for telling the truth 'to protect a right' is stronger in Bruce's case, since it involves a higher level of trust between Bruce and Mark. Second, the shared reason for not telling the truth 'to produce benefit' is weaker in Bruce's case, since Bruce's potential profit will not make much difference to his overall financial situation. Third, Stephanie has the reason to not tell the truth to strive for a greater good for the citizenry. Finally, Bruce's motivations for not telling the truth, unlike Stephanie's, appear to be purely selfish. This increases the onus on Bruce to tell the truth.

**Figure 1:** TRUTH-TELLER's Output Comparing Stephanie's and Bruce's Cases

Before tracing the algorithm, let us summarize what TRUTH-TELLER has done. The program starts by determining the degree of similarity between the two episodes. The degree of similarity provides what we refer to as the "comparison context" and dictates the overall structure of the comparison text. Notice that the program deems the cases in Figure 1 to be "similar dilemmas." It determines this because there are similar types of reasons for both telling the truth and not telling the truth. Similar reasons share ancestors in the reasons hierarchy. The episodes also display similarity regarding the relationships between the actors. The experimenter/subject relationship between Stephanie and the experiment subjects and the sibling relationship between Bruce and Mark both involve a high degree of responsibility and duty. This information is culled from the relations hierarchy.

Because these particular cases comprise a comparison context of similarity, the program starts by focusing on the similar features of the cases (paragraphs 1 to 3). It is worth noting, that the most salient feature is sometimes a difference rather than a similarity. For instance, in a comparison context in which only one case involves life threatening circumstances, the comparison focus shifts toward the critical distinction, i.e., the life threatening circumstance.

To distinguish the two cases the program next attempts to draw attention to the relative strength of telling the truth in each of the cases (paragraphs 4 to 6). Again, the comparison context determines this tactic; the cases have been determined to be similar and now they must be differentiated. In this context it is common to argue either (a) that one case clearly has a stronger basis for telling the truth or (b) the relative merits of telling the truth in each case. The program determines that neither Stephanie nor Bruce has a stronger case for truth telling and thus argues the relative merits of telling the truth in each case (paragraphs 5 and 6). These arguments are essentially constructed from (a) reasons that are stronger in one case than the other, (b) reasons that exist only in one case, and (c) actions that are better supported in one case than the other.

We now turn to a trace of TT's algorithm. The program starts by accepting semantic representations of each of the cases. The representation of a case is an interpretation of the story that describes it and is manually constructed. Figure 2 depicts the semantic representations of the Stephanie and Bruce cases. In Stephanie's case Stephanie is the truth teller, since it is she who is confronted with the decision to tell the truth or not. The experiment subjects will hear the truth, should Stephanie decide to divulge it, and thus are the truth receivers in this episode. Finally, the citizenry and the scientific community are affected others, since they would be affected by any truth telling disclosure (i.e., they stand to benefit in some way should the experiment result in an important scientific finding).

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

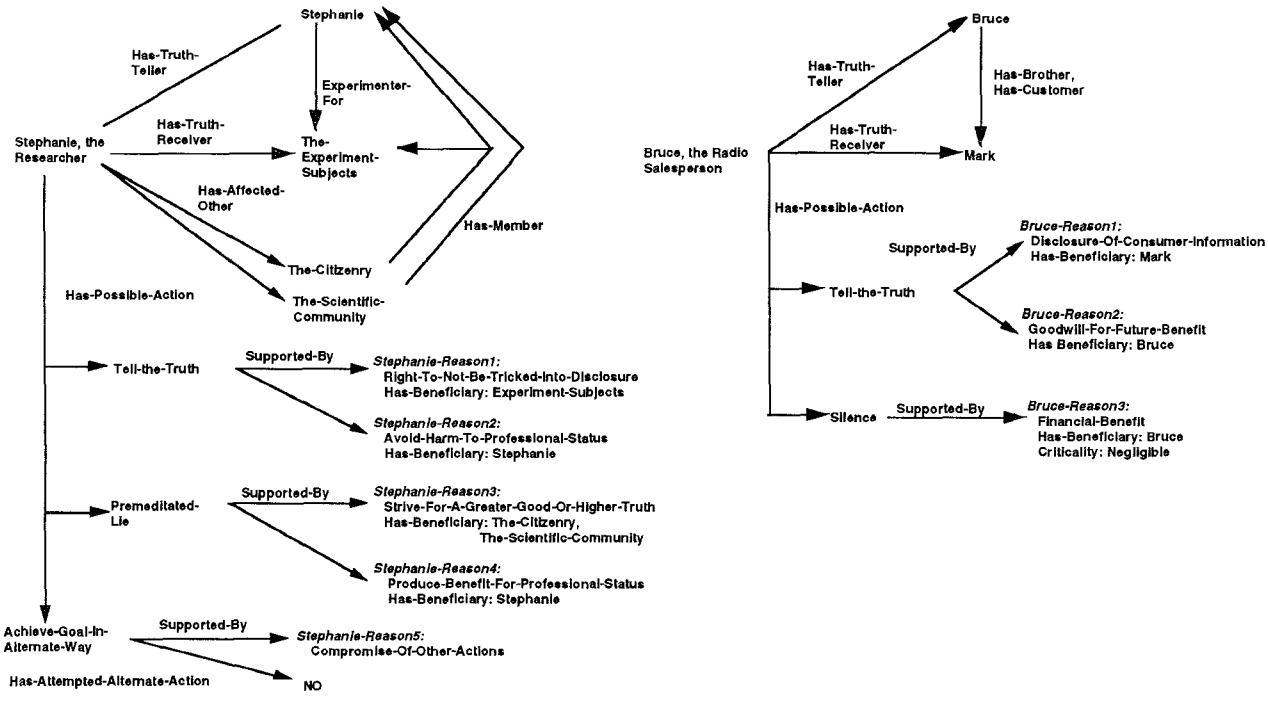


**Figure 2:** Representation of Stephanie's Case (left) and Bruce's Case (right)

of not telling the truth, for instance, telling a lie or keeping silent (i.e., not disclosing information). In Stephanie's case, the choice is between telling the truth about the intent of the experiment or deceiving the subjects by telling a premeditated lie. The lie is premeditated since, supposedly, Stephanie would know the intent of her own experiment and would therefore have the opportunity to reflect on the decision to lie or not. Stephanie also has an alternative action she could take before deciding whether or not to lie to the experiment subjects: she could pursue an alternative experimental design that would not require deceit.

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 Stephanie's telling the truth is to protect the right of the experiment subjects to not be tricked into the disclosure of information. A rationale for Stephanie telling the lie is to provide a "greater good" for the scientific community and the citizenry at large, i.e., the experiment may result in some general benefit for many people.

It is important to note that the semantic representations are interpretations of the stories, created manually as a knowledge engineering task. We manually assigned reasons as supporting various actions. We also manually added alternative actions (e.g., the possibility of Stephanie's changing the experimental design, which is not explicitly stated in the story). In the current work, we focus on the contribution that the marshaling techniques and other means for reasoning about reasons make to the program's ability to compare cases already represented with reasons and actions. We recognize the desirability of having the program infer reasons and actions based on past cases and hope to address that in future work.

The underlined portion of the comparison text of Figure 1 is part of the argument for Bruce having a more compelling case to tell the truth than Stephanie. A portion of the Stephanie and Bruce semantic networks and the comparison between these portions (Figure 3) are directly responsible for this text. The following paragraphs explain how this diagram and the four phases of the program led to the underlined comparison text.

Given the input representations, TRUTH-TELLER reasons about reasons by aligning, qualifying, and marshaling the semantic networks. First, the Alignment phase is initiated. The dashed lines in Figure 3 depict alignments between the Stephanie and Bruce representations. First, Stephanie's reason involving a "greater good" is determined to have no counterpart in the Bruce representation (i.e., it is a clear distinction); thus it is "misaligned" and labelled as a reason distinction. Second, the premeditated lie for Stephanie and silence for Bruce — possible actions for each case — are aligned with one another because they abstractly match as a "do not tell the truth" action. Finally, Stephanie and Bruce have reasons that abstractly match and are thus aligned with one another (i.e., Stephanie may benefit professionally by lying to her subjects and conducting the experiment, while Bruce may benefit financially by withholding the truth from his brother). These reasons are not identical; however, they match in the reason hierarchy at the level of "producing benefit." Reasons abstractly match if they share a common ancestor, up to and including the level of Bok's general reason types, i.e.,

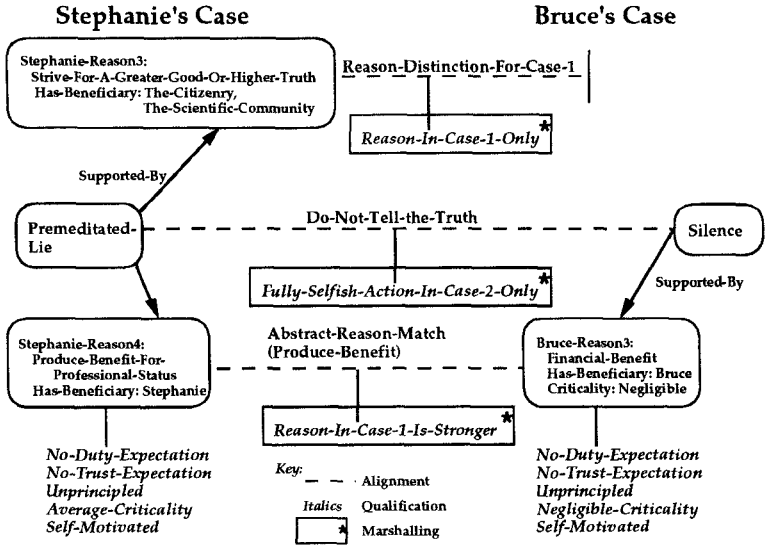


**Figure 3:** Comparison after Alignment, Qualification, and Marshaling

fairness, veracity, producing of benefit, and avoiding harm.

The program next commences the Qualification phase. The italicized text in Figure 3 represents the qualifications that are applied to the comparison. During this phase individual objects and alignments between objects are qualified or "tagged" based on finer-grained knowledge. The first step is to qualify the individual objects. For instance, Stephanie's reason to produce benefit (i.e., to improve her professional status) is tagged (1) as having no duty or trust expectations, since it only involves herself, (2) as being unprincipled, since no ethical principle supports the reason, (3) as having average criticality, since no comment was made in the story about any critical consequences, and (4) as being self-motivated, since the reason is clearly selfish. Bruce's aligned reason (i.e., financial benefit) is tagged likewise, with the exception that its criticality is labelled as negligible, since the story states that the financial benefit to Bruce is not important.

The second step of Qualification is to qualify alignments. 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 1. Second, the alignment between Stephanie's premeditated lie and Bruce's silence is tagged as being fully selfish, and thus weaker, on Bruce's side, since Bruce has only a selfish reason for withholding the truth, while Stephanie has one rationale, the "greater good" reason, that is not selfish. Finally, the abstract reason match between Stephanie's possible professional gain and Bruce's financial benefit is tagged as stronger for Stephanie. This is the case because of the superiority of average criticality over negligible criticality.

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 Stephanie/Bruce 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 then 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. As another example, if a comparison is assigned to the Incomparable-Dilemmas/Criti-

cality-Difference, the program marshals the reasons that have greater criticality as the comparison focus, and then marshals, as distinguishing information, the contrasting, less critical reasons of the less critical case.

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 an asterisk. Only the marshaled distinguishing information is depicted in the diagram, corresponding to the underlined text in Figure 1. 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 abstract reason match is a similarity, but qualification has also revealed it as a distinction. Note that the qualifier *Reason-In-Case-1-Is-Stronger* strengthens the case for Bruce telling the truth relative to Stephanie, since Bruce's reason is less compelling for not telling the truth. This marshaled data corresponds to the sentence in the comparison text beginning "Second, ..." Stephanie's reason for not telling the truth to attain a "greater good" is also marshaled as a strength of Bruce's case relative to Stephanie's, because it is a misaligned reason distinction (i.e., it provides justification for Stephanie to not tell the truth that is unshared by Bruce). This corresponds to the sentence "Third, ..." Finally, the program marshals the qualifier, *Fully-Selfish-Action-In-Case-2-Only*, from the matching actions in Figure 3. This also supports Bruce's case relative to Stephanie's, since it shows a weakness for not telling the truth that exists for Bruce but not Stephanie. This final marshaled data corresponds to the text in beginning "Finally, ..."

Figure 4 summarizes the ways that TT reasons about reasons. The Alignment phase employs methods 1 through 4. The Qualification phase employs methods 5 and 6. The Marshalling phase employs 7. Notice that our example has illustrated most of the reasoning techniques in the first two phases of the program: 3 and 4 (i.e., the abstract match and the distinction), 5 (i.e., the qualifications on the abstractly matched reasons), 6 (i.e., the comparison of the abstractly matched reasons), and 2 (i.e., the "fully selfish" comparison of the do-not-tell-the-truth actions). The comparison in its entirety employed all of the techniques.

1. **Elicit principles underlying reasons and classify reasons individually:** Reason Hierarchy follows links from reason type to principles. Classify reasons as principled, self-motivated, or altruistic.
2. **Classify reasons in the aggregate:** Note if all reasons supporting an action are principled or unprincipled, altruistic or self-motivated.
3. **Match reasons:** Identify reasons for a particular action shared by cases and reasons not shared. Matches may be exact or abstract, based on a hierarchy of reasons and principles. Also, note exceptional reasons in one or both cases or reasons distinct to one case.
4. **Map reason configurations:** Mark shared configurations of reasons such as shared dilemmas (i.e., similar opposing reasons).
5. **Qualify reasons by:** (a) criticality (what happens if action is not taken?), (b) whether altruistic or not, (c) whether principled or not, (d) participants' roles and relationships (e.g. trust, duty), (e) existence of untried alternative actions, (f) how others are affected by action, (g) comparing actions (Lie vs. Silence, Premeditated vs. Unpremeditated)
6. **Compare overall strength of reasons:** Use the qualifiers to decide whether one reason is "stronger" than another.
7. **Marshaling reasons:** Select, collect reasons to emphasize based on the overall similarity of cases, nature of the reason mapping and qualifications on reasons.

**Figure 4:** TRUTH-TELLER's Techniques for Reasoning about Reasons

The final phase of the program is the Interpretation phase. This phase generates the natural language depicted in Figure 1 by traversing subgraphs of an augmented transition network (McKeown, 1985). Each of the five comparison contexts 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. The Interpretation phase then translates the rhetorical predicates into the sentences that comprise the comparison text of Figure 1.

To summarize, the extended example shows how TRUTH-TELLER uses its 4-phase algorithm to generate context sensitive and marshaled case comparisons. 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 program marshals the reasons and qualifications by recognizing the context represented by a pair of cases. The comparison context dictates the strategy the program employs to marshal the relevant similarities and differences. Finally, the Interpretation phase accepts the marshaled information and generates a comparison text.

## 5. The Evaluation

Our goal was to obtain some assurance that TRUTH-TELLER's marshaling techniques and other techniques for reasoning about reasons 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.

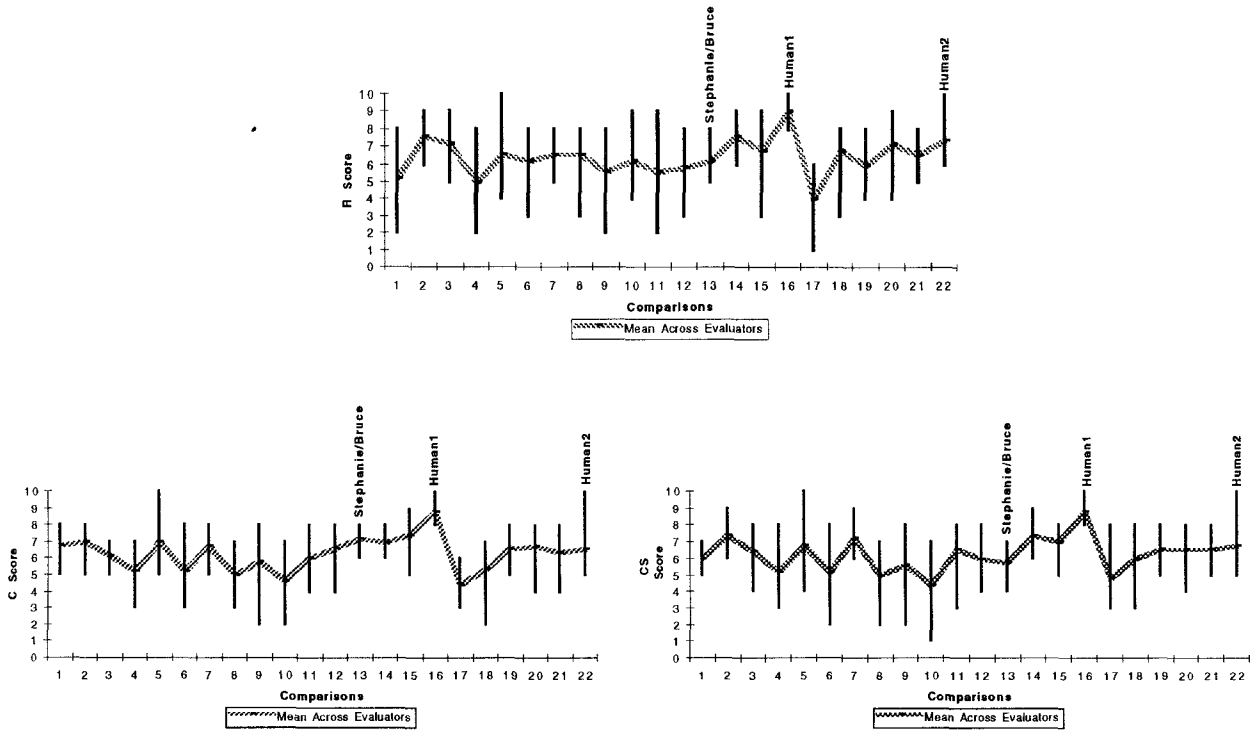


**Figure 5:** Max, Min, and Mean Values for R (top), C (left), and CS (right) Scores in Experiment #1

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 from three classes described below. The comparison texts were similar to and included the one in Figure 1. We also added two comparison texts generated by humans (a medical ethics graduate student and a law school professor). The experts were advised that the texts had been generated by humans or a computer program, but they were not told which texts or how many texts were generated by which. 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 twenty case pairs presented to TRUTH-TELLER were selected randomly as follows: (1) two cases selected at random from the *training set* (total of 5). (2) one case selected at random from the training set and one case selected at random from the *test set* (total of 5). (3) two cases selected at random from the test set (total of 10). The training set comprised twenty-three of the fifty-one cases in TRUTH-TELLER's case base; these cases were used to develop the program. The other twenty-eight cases served as the test set for the evaluation; these were used sparingly (or, in most cases, not at all) to develop the program.

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.

The mean scores for the Stephanie/Bruce comparison, number 13, were R = 6.2; C = 7.2; CS = 5.8. 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 repeatedly compared to a different second case (i.e., a One-To-Many comparison). 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 all evaluators 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.

## 6. 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. Given the instruction to "evaluate the comparisons as you would evaluate short answers written by college undergraduates," we are encouraged by the grades assigned. 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. We recognize, however, that such an experimental design provides greater assurance of the quality of any results and will employ it for a future summative 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 slightly higher scores in the second part of the experiment are due in part to the fact that it would have been easier for the evaluators to recognize TT's sensitivity to context in the one-to-many part of the experiment than in the first part. Upon recognizing the difficulty of context sensitive comparisons and TT's ability to tackle them, the evaluators tended, we believe, to grade the program higher on all dimensions. In fact, as was noted, the context sensitivity dimension graded higher than the other two dimensions in this part of the experiment.

Our primary mechanism for improving the TRUTH-TELLER model will be to respond to criticisms made by the evaluators. For instance, several evaluators questioned TRUTH-TELLER's lack of hypothetical analysis; the program makes immutable assumptions about reasons, actions, and actors, ignoring alternative interpretations. Addressing this would require a program imbued with a more elaborate, flexible representation; we have thought of using hypothetical variations along factors. Another repeated criticism involved abstract reason matches; there were a number of occasions in which an abstract match was questioned. This is probably due partly to disagreements about the structure of the reason hierarchy and partly to the level in the hierarchy in which reasons can be said to "match." For instance, one evaluator protested that "avoid emotional distress" and "avoid a reprimand" are "not at all the same," yet these abstractly matched as "avoid harm" reasons in one of TT's comparisons. It may be that humans, when reflecting on ethical dilemmas, typically think more in terms of exact matches of reasons or principles. We need to determine under what circumstances an abstract match is important and how far up the reason hierarchy TT should search for matches. Finally, aggregate reasons were questioned to a moderate degree. This was due, at least in part, to the fact that aggregate reasons are hard for the program to explain, since they require reference to other parts of the text. However, we have also considered that the calculus of support for actions and comparison of actions may be more complex than focusing simply on whether an action is fully supported by altruistic or principled reasons. For instance, one could argue that an action that is mostly altruistic, but happens to lead fortuitously to a selfish side benefit, is as good as a fully altruistic action. We intend to experiment with different levels of support (e.g., all altruistic but one) and present the results to experts for further evaluation.

In conclusion, the evaluation encourages us that TRUTH-TELLER makes mostly reasonable comparisons (although clearly not as sophisticated as humans and clearly requiring some improvements). Further, the program can make comparisons over a range of cases and displays some sensitivity to comparison context. We believe that the evaluation has shown that TT's AI CBR knowledge representation, marshaling process, and other techniques for reasoning about reasons provide a good start at developing a comparative evaluation model in the truth telling domain. We intend to improve TT's knowledge representation, based on the feedback from the formative evaluation, to work on developing a CEM like those proposed by Strong and Jonsen/Toulmin for ethical domains and to explore adapting it to legal domains. Some features of the program are readily adaptable to the legal domain: its ability to elicit principles underlying reasons, identify shared dilemmas, qualify reasons based on criticality and the participants' roles, relationships and interests, and marshal reasons.

## References


Aleven, V. and Ashley, K. D. (1994). An Instructional Environment for Practicing Argumentation Skills; In the Proceedings of *AAAI-94*, pages 485-492.

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

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.

Augustine (edited 1952). "Lying" and "Against Lying." In *Treatises on Various Subjects*, vols 14, 16. Edited by R. J. Deferrari, Catholic University of America Press.

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

Berman, D. and Hafner, C. (1993). Representing Teleological Structure in Case-Based Legal Reasoning: The Missing Link. In *Fourth International Conference on Artificial Intelligence and Law*, pages 50-59, Vrie University, Amsterdam. ACM Press: New York.

Bok, S. (1989). *Lying: Moral Choice in Public and Private Life.* Random House, Inc. Vintage Books, New York.

Branting, K. L. (1991). Building Explanations From Rules and Structured Cases. In the *Journal of Man-Machine Studies.* 34, pages 797-837.

Dworkin, R. (1977). *Taking Rights Seriously.* Cambridge, MA: Harvard University Press.

Frankena, W. K. (1973). *Ethics.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.

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

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

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

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.

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

Sidgwick, H. (1907). Classification of Duties - Veracity. In *The Methods of Ethics.* 7th edition London: MacMillan & Co.

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.